

EDUARDO SANTOS EBERHARDT

**PROCESSO DE AQUISIÇÃO, MANIPULAÇÃO E
DISTRIBUIÇÃO AUTOMATIZADA DE INFORMAÇÕES:
CASO APLICADO A UMA REDE DE FRANQUIAS**

Trabalho de Conclusão de Curso
submetido ao Departamento de
Engenharia Elétrica e Eletrônica da
Universidade Federal de Santa
Catarina para a obtenção do título de
Bacharel em Engenharia Elétrica.
Orientador: Prof. Dr. Mauricio
Valencia Ferreira da Luz.

Florianópolis
2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Eberhardt, Eduardo Santos
PROCESSO DE AQUISIÇÃO, MANIPULAÇÃO E
DISTRIBUIÇÃO AUTOMATIZADA DE INFORMAÇÕES. : CASO
APLICADO A UMA REDE DE FRANQUIAS / Eduardo Santos
Eberhardt ; orientador, Mauricio Valencia Ferreira
da Luz, 2018.
83 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro
Tecnológico, Graduação em Engenharia Elétrica,
Florianópolis, 2018.

Inclui referências.

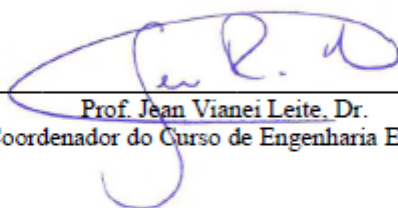
1. Engenharia Elétrica. 2. Extração de dados. 3.
Distribuição Automatizada de Informações. 4. Business
Intelligence. 5. BI. I. Valencia Ferreira da Luz,
Mauricio . II. Universidade Federal de Santa
Catarina. Graduação em Engenharia Elétrica. III.
Título.

Eduardo Santos Eberhardt

**PROCESSO DE AQUISIÇÃO, MANIPULAÇÃO E
DISTRIBUIÇÃO AUTOMATIZADA DE INFORMAÇÕES:
CASO APLICADO A UMA REDE DE FRANQUIAS**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Engenheiro Eletricista”, e aprovado em sua forma final pelo Programa de Graduação em Engenharia Elétrica.

Florianópolis, 04 de julho de 2018.



Prof. Jean Viane Leite, Dr.
Coordenador do Curso de Engenharia Elétrica

Banca Examinadora:



Prof. Mauricio V. Ferreira da Luz, Dr.
Professor da UFSC - Orientador



Prof. Walter Pereira Carpes Junior, Dr.
Professor da UFSC



Eng. Ricardo de Araújo Elias, M.Sc.
Doutorando do GRUCAD/UFSC

Este trabalho é dedicado a todos os
que fizeram parte da minha caminhada
até aqui.

AGRADECIMENTOS

Agradeço a todos que de alguma forma contribuíram para minha formação, tanto pessoal como profissional.

Pais, irmãos e família, muito obrigado por todo o amor, carinho e apoio incondicional que vocês me dão. Obrigado por fazerem tudo para o meu bem. Sou muito feliz e privilegiado por ter vocês.

Bons professores, obrigado por todos os ensinamentos repassados ao longo destes anos. Obrigado por mesmo apesar de todas as condições da educação no nosso país, darem o sangue por uma missão tão nobre que é repassar todo o conhecimento adquirido. Vocês são heróis!

Professor Maurício, primeiramente, obrigado por ser meu orientador e mentor. É uma honra imensa para mim. Obrigado por desde o início da minha graduação, preocupar-se em além de ensinar o conteúdo, nos fazer enxergar que a vida não espera e que se queremos algo, devemos nos esforçar e fazer acontecer.

Júlia, obrigado pela parceria, carinho, paciência e incentivo. Teu apoio e motivação foram fundamentais.

Amigos, obrigado por todos os momentos e lições que aprendemos juntos ao longo dessa jornada. Sem vocês a vida não teria graça.

Clientes e colegas, obrigado por confiarem em mim e por toda a troca de conhecimentos.

“All our dreams can come true if we have the courage to pursue them.” (Walt Disney)

RESUMO

No cenário atual do mercado, a tomada de decisão baseada em dados é cada vez mais uma necessidade. Através da tomada de decisão baseada em dados pode-se maximizar o número de decisões acertadas e minimizar as perdas, aumentando assim a competitividade das empresas. Este trabalho relata todas as etapas do processo de implementação de um projeto cujo principal objetivo é aumentar a receita de uma rede de franquias, através da distribuição tratada e automatizada das informações relevantes para tal. Inicialmente, são apresentadas as técnicas existentes para que seja possível implementar-se o projeto. Em seguida, é demonstrado o processo de definição do modelo a ser implementado. Na sequência, é mostrada como se deu a implementação do projeto. Finalmente, são apresentados os resultados obtidos e a discussão do trabalho.

Palavras-chave: Extração de dados. Distribuição Automatizada de Informações. Business Intelligence. BI. ETL.

ABSTRACT

Contemporary times demand knowledge of how to use data and technology to make decisions. Through data based decisions one can maximize the number of right decisions and minimize losses, improving the efficiency of companies. This article describes all the steps taken to implement a Business Intelligence project, in which the main goal is to increase revenues of a franchise player. The goal is meant to be reached through the automate distribution of treated and relevant information. First, it is presented the technical background and theory needed to understand the article. Then, it is shown how the project itself was implemented. The closure brings the results and discussion about the project.

Keywords: BI. ETL. Business Intelligence. Automate Data Distribution. Data Extraction.

LISTA DE FIGURAS

Figura 1- Descrição macro do problema proposto para o projeto.	25
Figura 2 - Exemplo de gráfico em QlikView.	30
Figura 3 – Exemplo de painel (ou dashboard) em QlikView.	31
Figura 4 – Estrutura básica de um banco de dados.	33
Figura 5 – Exemplo de chave de ligação entre tabelas de pedidos e vendedores.	35
Figura 6 – Fluxo típico de Conexão ODBC.	36
Figura 7 – Áreas contempladas pelo sistema Mavericks, ERP da empresa cliente.	43
Figura 8 – Exemplo de tela de vendas de um ERP.	44
Figura 9 – Modelo Estrela.	47
Figura 10 – Modelo Floco de Neve.	48
Figura 11 – Comparação entre modelos de dados.	48
Figura 12 – Matriz de decisão binária do modelo de dados a ser adotado.	50
Figura 13 – Trecho do código da extração do Mavericks (dados sigilosos foram ocultados).	51
Figura 14 – Trecho do código da aplicação de transformação dos dados.	53
Figura 15 – Trecho do código da aplicação de carga dos dados.	54
Figura 16 – Fluxograma de atualização dos dados configurado no QMC (Qlik Management Console).	55
Figura 17 – Hierarquia de gestão das lojas.	56
Figura 18 – Painel de carro, mostrando principais indicadores.	59
Figura 19 – Níveis de maturidade em data analytics e respectivos processos manuais e automatizados.	60
Figura 20 – Exemplo da função Sum, disponível para uso em QlikView.	62
Figura 21 – Exemplo funções combinadas para gerar indicador de previsão de vendas.	62
Figura 22 – Estratégia de distribuição automatizada de relatórios através do uso do Nprinting.	63
Figura 23 – Conexão do bloco de fontes de dados do Nprinting ao QlikView.	64
Figura 24 – Criação de relatório em Excel através do uso do bloco de relatórios do Nprinting.	65
Figura 25 – Configuração do e-mail de envio dos relatórios.	65

Figura 26 – Estimativa de tempo necessário para executar manualmente, para uma loja, o processo automatizado no projeto.....	67
Figura 27 – Estimativa de tempo necessário para executar manualmente o processo automatizado no projeto para todas as 151 lojas.....	68
Figura 28 – Tempo de execução das etapas do processo automatizado para todas as 151 lojas.....	70
Figura 29 – Tempo de execução das etapas de extração, integração e modelagem dos dados do processo automatizado para todas as 151 lojas.....	71
Figura 30 – Início do processo automatizado de geração dos indicadores, dos alertas e do menu de ações.....	72
Figura 31 – Fim do processo automatizado de geração dos indicadores, dos alertas e do menu de ações.....	72
Figura 32 – Início do processo automatizado de distribuição por e-mail dos relatórios.....	72
Figura 33 – Fim do processo automatizado de distribuição por e-mail dos relatórios.....	72
Figura 34 – Evolução do % VNPE no ano de implantação do projeto.....	73

LISTA DE TABELAS

Tabela 1 – Exemplo de registro de venda de um ERP.....	44
Tabela 2 – Ganho de eficiência obtido com automatização do processo desenvolvido no projeto.	70
Tabela 3 – Estimativa de diferença de VNPE obtida através da implementação do projeto.	74
Tabela 4 – Crescimento do indicador de VNPE obtido através da implementação do projeto.	75

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

B2C - Business to Client
BI - Business Intelligence
CAE - Carteira Ativa de Especificadores
CC - Carteira de Clientes
DCL - Data Control Language
DDL - Data Definition Language
DML - Data Manipulation Language
DQL - Data Query Language
ERP - Enterprise Resource Planning
ETL - Extract, Transform, Load
FAT - Faturamento
FTP - File Transfer Protocol
GB - Gigabyte
HTML - HyperText Markup Language
Hz - Herz
IP - Internet Protocol
KPI - Key Performance Indicator
ND - Número de Dias
NVE - Número de Vendas para Especificadores
ODBC - Open Database Connectivity
PDF - Portable Document Format
PMC - Período Médio entre Compras
QMC - QlikView Management Console
QVD - QlikView Data
Rad - Radiano
RAM - Random Access Memory
RPM - Rotações por Minuto
SGP - Sistema de Gestão Empresarial
SQL - Structured Query Language
TI - Tecnologia da Informação
V - Volt
VMVE - Valor Médio das Vendas para Especificadores
VNP - Venda na Ponta
VNPC - Venda na Ponta para Clientes
VNPE - Venda na Ponta para Especificadores

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Descrição geral do problema	24
1.2	Objetivos	25
2	REVISÃO BIBLIOGRÁFICA.....	29
2.1	<i>Business Intelligence</i>	29
2.1.1	QlikView	29
2.1.2	Nprinting.....	31
2.2	Bancos de dados	33
2.2.1	SQL.....	34
2.2.2	Chaves de ligação	34
2.3	Transação de dados entre sistemas	36
2.3.1	Conexão ODBC	36
2.3.2	FTP	37
2.4	Modelagem de dados.....	38
2.4.1	Tabela fato	38
2.4.2	Tabelas de Dimensões	39
3	MATERIAIS E MÉTODOS.....	41
3.1	ETL – Aquisição e tratamento dos dados	41
3.1.1	Mapeamento das fontes de dados disponíveis.....	42
3.1.1.1	Mavericks.....	42
3.1.1.2	Rabat	45
3.1.1.3	Padang.....	45
3.1.2	Estudo e comparação entre modelos de dados	46
3.1.3	Modelo de dados adotado	49
3.1.4	Extração dos dados	50
3.1.5	Transformação dos dados.....	52
3.1.6	Carga dos dados	53
3.1.7	Automatização do processo de ETL	54

3.2	Modelagem de negócio.....	56
3.2.1	Levantamento de informações e entrevistas.....	56
3.2.2	Interpretação das entrevistas.....	57
3.2.3	Definição dos indicadores (<i>KPIs</i>).....	57
3.2.4	Análises prescritiva.....	59
3.2.4.1	Alertas dos KPIs.....	60
3.2.4.2	Menu de ações	61
3.3	Implantação prática do modelo	61
3.3.1	Programação dos indicadores.....	61
3.3.2	Programação do Nprinting	63
3.3.2.1	Criação dos relatórios.....	64
3.3.2.2	Distribuição automatizada dos relatórios	65
3.4	Validação e ajustes.....	66
3.5	Lançamento	66
4	RESULTADOS	67
4.1	Eficiência e tempo de processamento.....	67
4.1.1	Estimativa de tempo para execução manual	67
4.1.2	Análise de eficiência e tempo de execução do processo automatizado	69
4.2	Alcance do projeto	72
4.3	Resultados financeiros	73
5	CONCLUSÃO.....	77
5.1	Sugestões para trabalhos futuros	77
6	REFERÊNCIAS.....	79

1 INTRODUÇÃO

Com o constante aumento do uso da tecnologia digital, o volume de dados está crescendo de maneira tão exponencial que nos últimos dois anos criaram-se mais dados do que em todo o restante da história prévia da raça humana [1].

Um dos motivos mais relevantes para este rápido crescimento é o fato de que se utilizam cada vez mais sistemas no dia a dia das pessoas. Aliado a isto, os sistemas contam com cada vez mais sensores, sejam estes físicos ou virtuais. Sensores são observadores de estados. Ou seja, cada vez mais são criados dados sobre o estado de máquinas, sistemas, interação intersistêmicas, interação das pessoas com sistemas, capturas de voz, imagens, texto e outros.

Como disse Aldous Huxley [2] “fatos não deixam de existir quando são ignorados”. Com isto, as organizações contam com um volume gigantesco de informações armazenadas em suas bases e lagos de dados (*data lake*).

Se bem utilizados, os dados auxiliam estas empresas a alcançar seus objetivos, ou até mesmo, defini-los [3]. Porém, em muitos casos, grande parte das informações disponíveis não é relevante ou não ajuda na gestão das empresas. Como sugere Clay Johnson em seu livro “A dieta da informação” [4] é essencial que não se trabalhe com informações que não agregam inteligência às análises. Desta forma, tem-se um ambiente com menos variáveis e fica mais fácil transformar dados brutos em informação de valor.

Do contrário, a tentativa de trabalhar e tomar decisões de negócio embasada em dados pode ter efeito contrário ao desejado e acabar com a eficiência de uma área ou de uma organização inteira.

Este trabalho descreve um projeto realizado pela empresa BIX Tecnologia para uma rede de franquias, a qual tinha necessidade de extrair valor das informações existentes. Esta rede de franquias é responsável pela frente de comercialização B2C (*business to clients*) de uma corporação mundial com vastíssimo volume de dados disponíveis, provenientes de diversas fontes.

O autor inicia o texto trazendo *background* teórico necessário para compreensão dos capítulos posteriores. Na sequência, apresenta aspectos técnicos referentes à modelagem de dados. Posteriormente, traz como foi o processo de definições das regras e necessidades de negócio da empresa. Então, aborda tópicos referentes ao desenvolvimento prático da geração e distribuição dos indicadores. Por fim, realiza análise dos resultados e discussão sobre o trabalho.

1.1 Descrição geral do problema

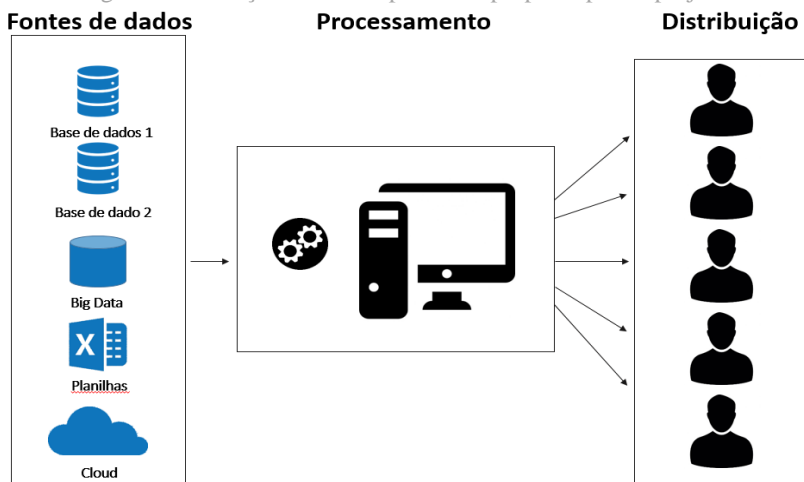
O presente trabalho descreve projeto de *Business Intelligence* realizado em uma rede de franquias com faturamento superior a R\$ 600 milhões por ano e mais de 150 lojas espalhadas em todo o território brasileiro. Apesar de ser uma empresa com grande faturamento, os gestores das áreas de *marketing* e comercial apostavam que era possível melhorar. Eles sabiam que possuíam muitas informações valiosas que eram subutilizadas e poderiam melhorar o faturamento da empresa.

Porém, não sabiam como e por onde começar. A empresa possuía muitas fontes de dados, a ponto de não saber nem exatamente quais fontes e quais dados eles possuíam.

Sendo assim, era necessário organizar as informações de maneira estratégica, processá-las e disponibilizá-las para os franqueados, para que os mesmos pudessem identificar as principais oportunidades para crescer em vendas e lucratividade.

Uma macro descrição do problema pode ser feita conforme figura 1, em 3 macro etapas. A primeira etapa é a integração de todas as fontes de dados disponíveis. A segunda etapa é processar as informações extraídas do sistema a fim de criar indicadores e relatórios que permitam ao usuário final a identificação das oportunidades de sua loja. Por fim, é realizada a distribuição das informações. Nesta etapa, são enviadas as informações específicas de cada loja para seus respectivos gestores.

Figura 1- Descrição macro do problema proposto para o projeto.



Fonte: Elaborado pelo autor (2018).

O projeto teve seis meses de duração e foi desenvolvido para as áreas de *marketing* e comercial.

1.2 Objetivos

O objetivo final do projeto é aumentar o faturamento da empresa. O faturamento da rede de franquias é dado pelo somatório das vendas para clientes finais com as vendas através de especificadores, conforme mostra a equação (1):

$$Fat = VNPC + VNPE \quad (1)$$

Onde:

Fat = Faturamento da empresa.

VNPC = Venda na ponta para clientes finais.

VNPE = Venda na ponta para especificadores.

Venda na ponta significa venda na loja. Especificadores diz respeito a um tipo de cliente desta rede de franquias, os quais recomendam os produtos destas lojas e intermediam as vendas para clientes finais.

O foco desta rede de franquias é realizar vendas para especificadores. No ano anterior ao ano de execução deste projeto, 70% do faturamento da empresa veio através de vendas para especificadores.

O valor de venda na ponta para especificadores é dado através de (2):

$$VNPE = NVE \cdot VMVE \quad (2)$$

Sendo:

$VNPE$ = Venda na ponta para especificadores.

NVE = Número de vendas realizadas para especificadores.

$VMVE$ = valor médio das vendas realizadas para especificadores.

NVE é dado por (3):

$$NVE = (CAE / PMC) ND \quad (3)$$

Sendo:

NVE = Número de vendas realizadas para especificadores.

CAE = Carteira ativa de especificadores. São considerados ativos, especificadores com compras nos últimos seis meses.

PMC = Período médio entre compras no período analisado.

ND = Número de dias no período analisado. Se o período analisado for um ano não bissexto, $ND = 365$.

Portanto, visando aumentar o faturamento através do aumento das vendas para especificadores ($VNPE$), definiu-se como objetivos do projeto:

1. Aumentar o % de vendas para especificadores (% $VNPE$).
2. Diminuir intervalo médio entre compras (PMC).
3. Aumentar o valor médio das vendas (VMV).

Para viabilizar os objetivos traçados, desejava-se implantar uma distribuição automatizada de relatórios para os franqueados. Para tal, os principais objetivos técnicos do projeto eram:

1. Mapear fontes de dados disponíveis.

2. Extrair e modelar os dados, integrando os dados de todas as fontes.
3. Propor melhor maneira de disponibilizar informações para os franqueados, visando maior lucratividade para rede de franquias.
4. Montar um processo 100% automatizado de extração, processamento e distribuição das informações.
5. Implantar o projeto em menos de nove meses.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta a descrição teórica dos métodos utilizados para a realização deste trabalho. Inicia-se apresentando sobre *Business Intelligence*. Em sequência são apresentadas noções de bancos de dados, transação de dados entre sistemas e finaliza abordando questões sobre modelagem de dados. Todos os pontos destes temas são apresentados de forma a auxiliar o desenvolvimento do projeto, apresentado no próximo capítulo.

2.1 *Business Intelligence*

O termo *Business Intelligence* (BI), inteligência de negócios, refere-se a uma ampla área de negócio das empresas. Envolve todos os aspectos referentes à coleta, organização, análise, compartilhamento, monitoramento e distribuição de informações que oferecem suporte a gestão de negócios.

“*Business Intelligence* é um conjunto de metodologias, processos, arquiteturas e tecnologias que transformam uma grande quantidade de dados não processados em informações úteis. Estas informações devem ser dispostas de forma a possibilitar maior efetividade nas visões estratégicas, táticas, operacionais, e também auxiliar a tomada de decisões.” [5]. Nasceu através da necessidade de automatizar a geração de relatórios que embasam a tomada de decisão nas empresas.

Business Intelligence pode servir para embasar tomada de decisões, monitorar a concorrência, alimentar pesquisas de marketing, acompanhamento de campanhas, e qualquer outra demanda estratégica da empresa que necessite de amparo técnico e de informações.

2.1.1 QlikView

O QlikView é um *software* voltado para o desenvolvimento de projetos de *Business Intelligence*. Oferece um pacote completo de soluções, passando por extração e integração de fontes de dados, modelagem das informações, administração de ambientes de acesso, distribuição automatizada das informações e acompanhamento de uso das aplicações criadas.

O QlikView foi desenvolvido pela empresa sueca QlikTech. É ferramenta pioneira no mercado de BI. Foi reconhecida pelo relatório do

Magic Quadrant, do Gartner [6] como ferramenta líder de mercado nos últimos oito anos.

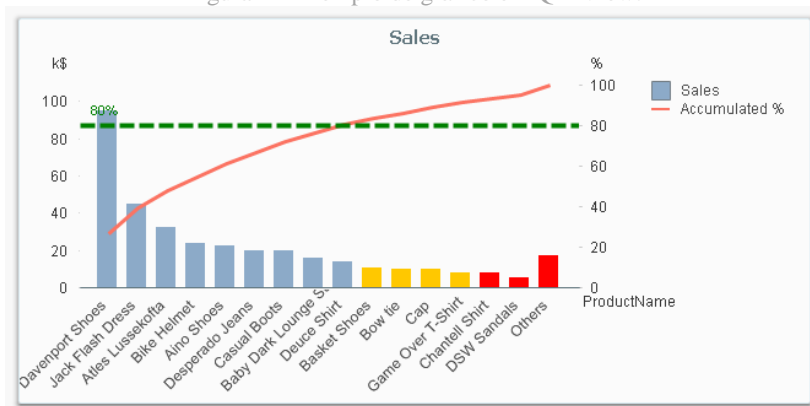
Pelo fato de ser uma ferramenta muito robusta e completa, a área de TI da empresa adquiriu licenças do QlikView e decidiu que o projeto seria desenvolvido com base na ferramenta.

O desenvolvimento é feito utilizando a linguagem de programação denominada pelo mesmo nome da ferramenta, QlikView. A linguagem permite programar rotinas referentes tanto à extração e modelagem de dados como a programação de indicadores. Os códigos referentes ao processo de extração e modelagem são denominados scripts de ETL. Já os códigos referentes ao processo de criação de indicadores são denominados scripts de expressões.

A linguagem QlikView possui funções pré-definidas. Estas funções pré-definidas têm por objetivo facilitar o desenvolvimento de indicadores.

Uma vez programadas as expressões que definem os indicadores, pode-se encapsular estas expressões em objetos da ferramenta. Para tal, definem-se as dimensões e expressões atreladas a cada objeto. Assim, cria-se um gráfico do QlikView, conforme mostra figura 2:

Figura 2 - Exemplo de gráfico em QlikView.

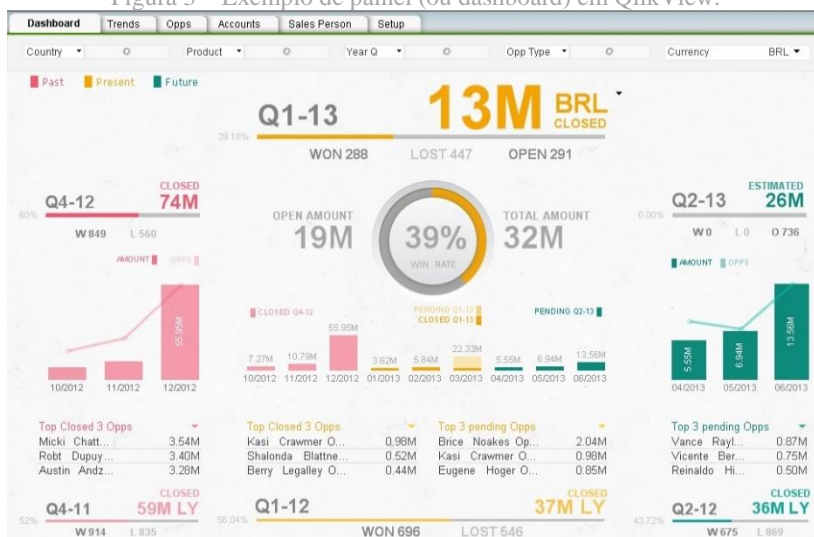


Fonte: Retirado de [7].

Após programar as expressões dos indicadores e criar respectivos gráficos com expressões e dimensões, diversos objetos podem ser unidos em uma mesma pasta do QlikView. Desta forma, é criado um

painel de gráficos e indicadores, também chamados de *dashboards*, conforme mostra figura 3.

Figura 3 – Exemplo de painel (ou dashboard) em QlikView.



Fonte: Retirado de [7].

Estes painéis são utilizados para realizar a análise das informações dentro do ambiente de acesso ao QlikView. O usuário pode filtrar os gráficos de maneira intuitiva, clicando ou selecionando os valores que deseja filtrar, de acordo com a análise de deseja realizar.

2.1.2 Nprinting

O Nprinting é uma ferramenta de distribuição automatizada de relatórios. Faz parte do pacote de softwares da QlikTech e funciona, exclusivamente, plugado às ferramentas de BI desenvolvidas pela QlikTech.

Por ser a solução recomendada pelo fabricante para realizar a distribuição de relatórios desenvolvidos em QlikView, adotou-se a ferramenta como método de distribuição dos relatórios para e-mails.

O processo de distribuição das informações geradas no QlikView depende da criação de relatórios em Nprinting.

Para tal, programam-se blocos de ações no Nprinting. Entre os blocos de ações, vale destacar os blocos:

- **Atualização:** Bloco em que se programa toda a parte referente à atualização das aplicações QlikView antes de se gerar os relatórios em Nprinting, garantindo que as informações distribuídas estejam atualizadas.
- **Contatos:** Bloco do Nprinting responsável por configurar quem serão os recipientes dos relatórios distribuídos. A ferramenta permite que o desenvolvedor configure a distribuição de informações de forma que ao gerar os relatórios, sejam filtradas informações referentes a cada um dos contatos. Isto é de extrema importância, pois permite que se realize o envio automatizado de relatórios com informações personalizadas para cada usuário final.
- **Relatórios:** Bloco em que se programam os relatórios propriamente ditos. É o bloco em que se realiza o link entre os objetos desenvolvidos em QlikView com os relatórios a serem distribuídos. A ferramenta permite gerar relatórios nos formatos Excel, HTML, Word, Power Point e PDF.
- **Mensagem:** Bloco em que se programa remetente, título, cabeçalho, texto, imagens e outras características do e-mail no qual o relatório será enviado.
- **Tarefas:** Bloco responsável por toda a parte de automatização dos outros blocos. Ou seja, neste bloco, é programada a execução do conteúdo dos outros blocos.
- **Fontes:** Bloco em que se define a origem das informações, que no caso devem ser aplicações de QlikView. Neste bloco, são realizadas as conexões necessárias entre as aplicações QlikView que contém os objetos que se utilizará nos relatórios e o Nprinting.

2.2 Bancos de dados

“A coisa mais valiosa de qualquer organização são seus dados. É onde está toda sua informação vital, como arquivos, pastas, propostas, base de dados de ERP, etc.” [8].

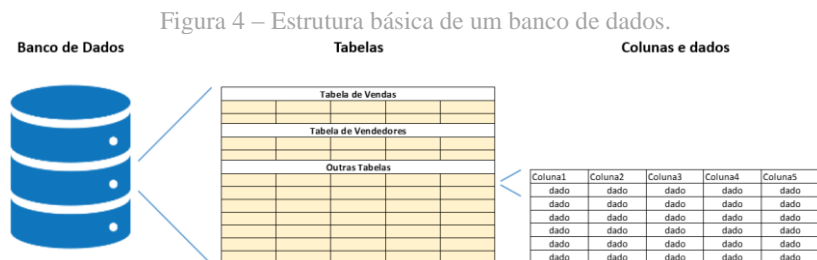
Conforme é afirmado em [8], os dados são muito valiosos para qualquer organização. Neste âmbito, não se fala de dados apenas como dados registrados em mídias digitais, mas também documentos de papel e outras formas de registros.

Estes dados carregam toda a verdade sobre uma empresa e sem eles, a empresa não possui nada registrado. Existem diversos exemplos de empresas que após o ataque de 11 de setembro de 2001, não conseguiram se reestabelecer por terem perdido por completo todos os seus bancos de dados, o que impossibilitou recuperação das informações da empresa.

Dada a importância de se guardar dados, é de suma importância que se realize o processo de armazenamento de dados da maneira mais organizada e otimizada o possível. Devido a esta necessidade, surgiram os bancos de dados.

Bancos de dados são estruturas, digitais, em que se armazenam dados brutos.

“Podemos entender por banco de dados qualquer sistema que reúna e mantenha organizada uma série de informações relacionadas a um determinado assunto em uma determinada ordem.” [9].



Fonte: Elaborado pelo autor (2018).

Conforme ilustra a figura 4, para armazenar um dado em um banco de dados, é necessário criar tabelas. Dentro das tabelas são criadas colunas, onde as informações são armazenadas. Para os dados da base de dados ficarem organizados, devem ser criadas uma ou mais tabelas para cada grupo de informações. Estas tabelas se relacionam entre si através de chaves, conforme explicado na subseção 2.2.2.

2.2.1 SQL

A linguagem de programação *Structured Query Language*, ou simplesmente SQL, é a base para utilização de bancos de dados relacionais.

A linguagem SQL é implementada através de comandos pré-definidos. Os comandos SQL são divididos em:

- DDL (*Data Definition Language*) ou Linguagem de definição de dados
- DML (*Data Manipulation Language*) ou Linguagem de manipulação de dados
- DQL (*Data Query Language*) ou Linguagem de recuperação de dados
- DCL (*Data Control Language*) ou Linguagem de controle de dados

Apesar de existirem diversos comandos SQL, através dos principais comandos desta linguagem (insert, delete, update e select) é realizada a maior parte das operações executadas em um banco de dados. Entre as tarefas executadas, destacam-se tarefas como:

- Criação das estruturas básicas de armazenamento, como campos e tabelas.
- Manutenção das estruturas, como alterações em nomes e outras propriedades.
- Extração de dados, através de *Queries*.

2.2.2 Chaves de ligação

Conforme já explicado neste capítulo, um banco de dados possui diversas tabelas em sua estrutura. Como ditam a literatura e as boas práticas de desenvolvimento referentes a banco de dados, cada tabela deve ser voltada a um assunto principal. Não se deve misturar mais de um assunto na mesma tabela do banco.

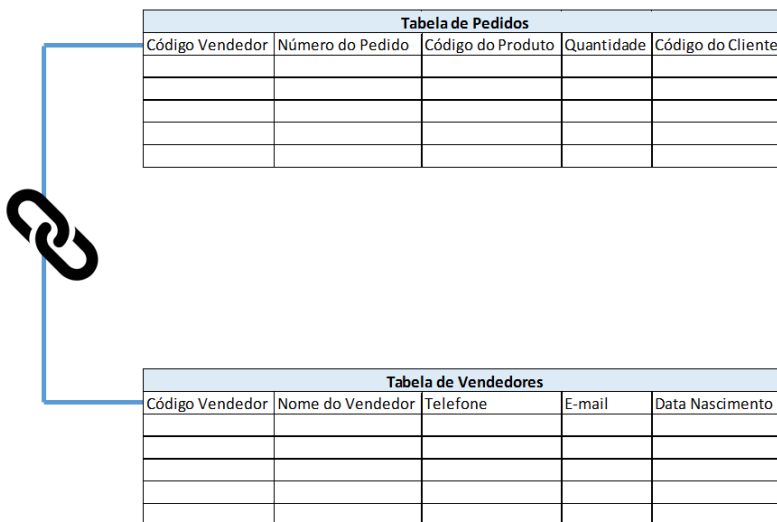
Por exemplo, supondo uma empresa que possui dados sobre vendas. Não se deve misturar dados de vendedores, clientes, pedidos e produtos em uma única tabela. Deve-se ter uma tabela para registrar informações do vendedor, uma para informações dos clientes, uma para informações dos pedidos e, por fim, uma para informações dos produtos.

Porém, na prática, os dados gerados são atrelados a diversas áreas de um negócio. Como no exemplo acima, um único pedido geraria dados para todas as tabelas citadas. Então, para conseguir guardam-se dados referentes a cada área do negócio em sua respectiva tabela, porém, permitir que seja possível interligar os dados entre as tabelas, utilizam-se as chaves de ligação entre tabelas.

Chaves de ligação entre tabelas são colunas das tabelas, que podem ou não ter o mesmo nome, que permitem a correlação entre os dados de duas ou mais tabelas [10]. Por exemplo, para o caso acima, haveria uma coluna que indicaria o código do vendedor tanto na tabela de pedidos como na tabela de vendedores.

Desta forma, seria possível saber os dados do vendedor que realizou a venda, interligando informações das duas tabelas. Analogamente ao processo descrito, pode-se aplicar este processo para as outras tabelas do banco de dados. Este processo pode ser entendido através da figura 5.

Figura 5 – Exemplo de chave de ligação entre tabelas de pedidos e vendedores.



Fonte: Elaborado pelo autor (2018).

2.3 Transação de dados entre sistemas

Como citado em 2.2, “A coisa mais importante de uma empresa são seus dados”. Mas pouco adianta possuir dados se não for possível acessá-los. Independentemente de qual maneira eles estão armazenados.

Sendo assim, é de extrema importância que se consiga garantir que os dados possam ser acessados pelos sistemas e pessoas que precisarem das informações.

Existem diversas maneiras de integrar (transacionar dados entre sistemas) os dados armazenados em sua origem primária à aplicação de destino final. É possível separar os métodos de integração em duas grandes abordagens:

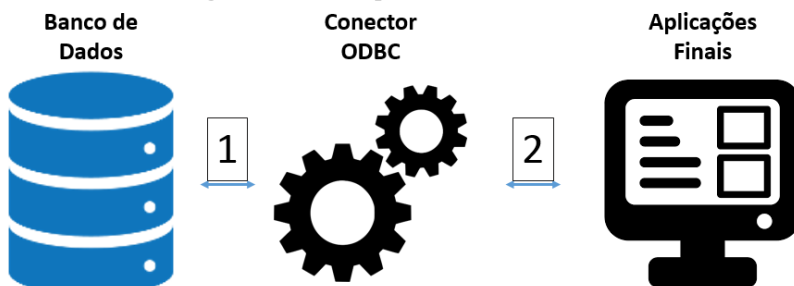
- Métodos de integração direta entre banco de dados e aplicação final.
- Métodos de integração com etapa de replicação dos dados entre banco de dados e aplicação final.

O objetivo desta seção é explicar quais os métodos de integração foram utilizados para garantir que todas as fontes de dados pudessem ser integradas.

2.3.1 Conexão ODBC

Uma das maneiras mais utilizadas para integrar as ferramentas de BI diretamente a bancos de dados é através de conexões do tipo *Open Database Connectivity*, ou simplesmente ODBC.

Figura 6 – Fluxo típico de Conexão ODBC.



Fonte: Elaborado pelo autor (2018).

Como demonstra a figura 6, um fluxo típico de conexão ODBC possui basicamente duas etapas, descritas abaixo:

- Etapa 1 – Conexão entre o conector ODBC e o banco de dados.

Esta etapa é responsável por realizar a conexão do conector com as tabelas do banco. De maneira simples, pode-se dizer que é a etapa responsável por “abrir as portas” para acesso ao banco de dados. Para tal, devem ser configurados parâmetros como usuário, senha, tabelas a serem acessadas e outros.

- Etapa 2 – Conexão entre conector ODBC e aplicações finais.

Esta etapa é responsável por ligar as aplicações finais ao banco de dados, através das conexões criadas pelo conector na etapa 1. De maneira simples, pode-se dizer que é a etapa responsável por extrair os dados das tabelas acessadas.

2.3.2 FTP

O termo FTP (*File Transfer Protocol*) refere-se a um protocolo de transferência de dados, como diz seu próprio nome se traduzido para o português.

É um repositório utilizado para armazenamento ou troca de dados entre sistemas. Funciona como uma pasta convencional de computadores, porém, é acessível de qualquer lugar com conexão à internet. Esta acessibilidade deve-se ao fato de que FTPs necessariamente precisam de um servidor que serve como diretório real da aplicação.

Para que a troca de dados intersistêmicos aconteça, um sistema deve escrever seus dados no FTP e o outro sistema deve ler estes dados. Sendo assim, diferentemente de conexões ODBC, serve como uma etapa de replicação dos dados armazenados em um banco de dados. É importante ressaltar que diferentemente dos bancos de dados, é possível armazenar qualquer tipo de arquivo dentro de repositórios FTP, sem um padrão pré-definido de tabelas ou colunas.

Para escrever ou adquirir dados de um FTP, é necessário estabelecer conexão com o mesmo. Para tal, pode-se acessar FTPs por meio de programas que funcionam como interfaces de acesso ou através de linhas de comando em terminais de sistemas. Independentemente da forma de acesso utilizada, são usados os parâmetros abaixo:

- Endereço: IP do servidor a ser acessado.
- Porta: parâmetro que sinaliza através de qual porta do servidor o acesso será feito. Para estabelecer a conexão utiliza-se como padrão a porta 21. Para realizar a escrita ou aquisição dos dados, utiliza-se como padrão a porta 20.
- Usuário: nome do usuário de acesso ao servidor. Somente usuários cadastrados e com acesso ao servidor são aceitos para estabelecer a conexão.
- Senha: senha do usuário passado como parâmetro.

Por ser tão flexível, ter um funcionamento simples e permitir fácil acesso para escrita e aquisição dos dados, possui uma grande vantagem em termos de velocidade de implementação de transferências de dados entre sistemas.

2.4 Modelagem de dados

“Se eu tivesse oito horas para cortar uma árvore, gastaria seis afiando meu machado” – Abraham Lincoln.

Esta frase ajuda a entender o papel da modelagem de dados em qualquer projeto de BI. Assim como é necessário uma boa escolha da ferramenta de BI, dos métodos de transações de dados e de diversos outros aspectos, uma má modelagem de dados por trás das camadas de visualização das informações podem significar o fracasso de um projeto. Sendo assim, é de extrema importância que analisar e realizar testes de eficiência com diferentes abordagens de modelagens de dados.

Isto ocorre, pois, como mencionado na introdução deste trabalho, em projetos de BI e *Big Data* pode-se trabalhar com um volume colossal de dados. Desta forma, qualquer tipo de perda de eficiência devido a modelagem de dados significa um altíssimo desperdício em custo computacional e performance das ferramentas.

Este subcapítulo demonstra como funcionam modelagens de dados, explicando sobre tabelas fatos e dimensões. É usada analogia com sistemas elétricos para facilitar a compreensão destes conceitos.

2.4.1 Tabela fato

A tabela central de um modelo de dados é chamada de tabela fato [11]. Poder ser compreendida como o coração dos dados de uma área de

negócio da empresa. É com base nos dados contidos nesta tabela que os indicadores são gerados.

O principal papel da tabela fato é armazenar os dados transacionais do sistema em análise. Como são gerados muitos dados transacionais, esta tende a ser a tabela com o maior número de linhas de todo o modelo de dados. Portanto, para evitar que a tabela fique muito pesada, é de extrema importância que não possua mais colunas do que o necessário. Para tal, as tabelas fatos são compostas de dois tipos de colunas. Estes são colunas de valores e colunas de chave de ligação [12].

As colunas de valores são as colunas que contém os dados sobre os eventos ocorridos. As colunas de chave de ligação contém códigos de ligação com as tabelas de dimensão e servem como identificador dos eventos ocorridos.

Em uma analogia com sistemas elétricos, a tabela fato gerada a partir de sensores de uma subestação de energia teria:

- Tensão (V), frequência (Hz) e fase (rad) como valores.
- Código da subestação e código da máquina como chaves de ligação.

2.4.2 Tabelas de Dimensões

Ao contrário da tabela fato, as tabelas de dimensões servem para permitir o detalhamento das informações transacionais. Armazenam dados descritivos. Como seus dados não são transacionais têm número de linhas limitados. Em geral, o número de colunas não é uma preocupação, pois quanto mais colunas uma tabela de dimensão possuir, mais detalhes sobre os fatos possui [13].

As ligações entre a tabela fato e as tabelas de dimensões são feitas através de chaves, conforme explicado em 2.2.2. Para cada chave de ligação da fato, existe pelo menos uma tabela de dimensão.

Tomando o exemplo apresentado da medição de uma subestação, o modelo de dados teria pelo menos duas tabelas de dimensão:

- Tabela de dimensão com as informações de subestações, como código, endereço, potência nominal, data de abertura, etc.
- Tabela de dimensão com informações das máquinas, como código, tipo de máquina, potência nominal, data da última manutenção, etc.

3 MATERIAIS E MÉTODOS

Este capítulo apresenta o desenvolvimento detalhado de aspectos práticos do projeto. Está dividido em quatro macro partes:

- ETL – Aquisição e tratamento dos dados;
- Aspectos do negócio e proposta de modelo;
- Programação do modelo proposto;
- Validação e lançamento.

3.1 ETL – Aquisição e tratamento dos dados

ETL é a sigla para Extração, Transformação e Carga (*Extract, Transform, Load*).

ETL refere-se a todo o fluxo dos dados, desde sua extração na base primária em que está armazenado, até sua utilização nas aplicações finais [14]. Nesta seção serão apresentados todos os assuntos referentes a estas etapas de aquisição, tratamento e carga dos dados.

Esta etapa é fundamental para o sucesso do projeto, pois os dados são toda a base para todo o restante deste. Uma modelagem robusta e bem feita de dados é essencial para que se possa escalar o projeto de protótipo para projeto em produção. Em etapa de protótipo, o volume de informações é controlado e limitado. Já em produção, o volume de informações é gigantesco e qualquer otimização representa uma diferença significativa. Fazendo uma analogia com sistemas elétricos, é como se estivéssemos tratando de uma grande usina de geração da energia. Como as ordens de grandeza são muito grandes, qualquer percentual de perda representa um valor absoluto de perda muito grande, portanto, não pode ser negligenciado.

Para explorar com clareza as etapas práticas desenvolvidas, as subseções apresentam-se divididas em:

- Mapeamento das fontes de dados disponíveis;
- Comparação entre modelos;
- Modelagem de dados adotada;
- Extração dos dados;
- Transformação dos dados;
- Carga dos dados;
- Automatização dos processos de ETL.

Ao final desta seção, tem-se a base de dados utilizados no projeto consolidada e é possível construir o produto final que agregará valor ao cliente.

3.1.1 Mapeamento das fontes de dados disponíveis

Como o desenvolvimento deste projeto contemplava a especificação do que precisava ser desenvolvido, a primeira tarefa a ser realizada era conhecer as possibilidades. Ou seja, conhecer as fontes de dados disponíveis para realização do projeto.

Desta forma, durante duas semanas foram analisados quais as fontes de dados disponíveis e suas respectivas características.

Basicamente, mapeou-se a possibilidade de extrair dados de três fontes de dados, as quais podiam integrar-se.

Nas próximas subseções são explicadas as principais características destas fontes e quais tipos de dados foram obtidos a partir de cada uma destas (os nomes foram alterados para manter dados sigilosos do projeto):

3.1.1.1 Mavericks.

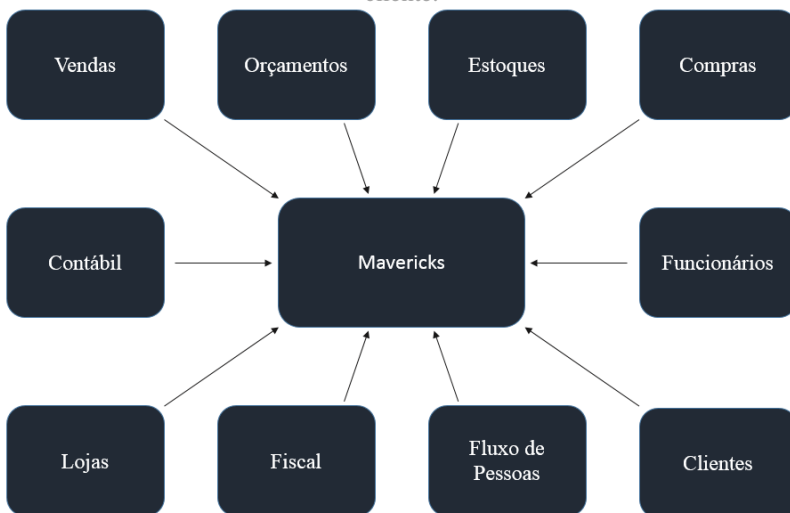
O *ERP (Enterprise Resource Planning)*, também chamado de Sistema de Gestão Empresarial (SGE, em português) é uma ferramenta corporativa cujo objetivo é controlar todas as informações transacionais e operacionais de uma empresa.

É uma ferramenta que deve ser capaz de gerar e armazenar dados sobre as mais diferentes áreas da empresa. Sua eficiência é dada através do fato de que em um *ERP*, todos os dados e processos são gerenciados em um único ambiente.

No caso específico da empresa em que foi aplicado o projeto, o *ERP* é chamado de Mavericks. Conforme mostra a figura 7 o sistema Mavericks contempla as áreas de vendas, orçamentos, estoques, compras, funcionários, cadastro de clientes, financeiro, lojas e fluxo de pessoas.

Todas as informações utilizadas no projeto referentes a estas áreas de negócio foram integradas a partir deste sistema.

Figura 7 – Áreas contempladas pelo sistema Mavericks, ERP da empresa cliente.



Fonte: Elaborado pelo autor (2018).

Para cada uma das áreas contempladas pelo Mavericks, existe uma tabela em um banco de dados MySQL. Estas tabelas são correlacionadas através de chaves primárias, conforme explicado na seção 2.2.2.

Cada funcionário da corporação possui um *login* e senha de acesso ao sistema, tendo acesso às funcionalidades de acordo com sua área de trabalho. A interação entre usuários e sistema é feita através de telas, com campos que devem ser preenchidos e então registrados. Na figura 8, mostra-se um exemplo de uma tela de vendas de um *ERP*.

Então, o sistema automaticamente escreve nas tabelas:

1. Que o estoque dos produtos comprados diminuiu nas respectivas quantidades.
2. Que o vendedor Daniel (código 2651) realizou a venda.
3. Que o cliente João (código 20169) realizou a compra.

3.1.1.2 Rabat

Trata-se de um sistema cujo objetivo é capturar informações públicas dos clientes e reuni-las em um único lugar de maneira simples e organizada. O sistema foi desenvolvido de maneira customizada para a empresa na qual o projeto foi realizado. Ele busca informações sobre cidadãos brasileiros de todo o território nacional, com enfoque em pessoas que possuam as características típicas dos clientes da empresa.

Para conseguir estas informações, Rabat utiliza técnicas de *scraping*. Esta técnica consiste em capturar e tabular informações disponibilizadas em outros sistemas [16]. Para tal, o algoritmo acessa páginas da internet, interpreta as informações disponibilizadas e grava-as em banco de dados próprio.

A ideia é que com os dados capturados por este sistema, seja possível identificar oportunidades de crescimento em vendas, realizar análises de público, potencial das regiões entre outros. Para facilitar a identificação de padrões de comportamento, busca-se o maior volume possível de informações.

3.1.1.3 Padang

Trata-se do site oficial da empresa para qual o projeto foi desenvolvido. Este site funciona também como portal para o programa de fidelização de clientes da companhia. Neste portal, os clientes têm acesso a diversos conteúdos gratuitos e podem trocar pontos por prêmios e benefícios. Os pontos são acumulados em compras e promoções realizadas nas lojas.

Padang possui duas fontes de dados que foram utilizadas neste projeto, informações cadastrais e informações sobre o programa de fidelidade. As informações cadastrais funcionam como uma fonte extra de informação sobre os clientes. Como as informações são digitadas ativamente pelos clientes com interesse em participar do programa de fidelidade, possui uma base mais qualificada em relação às informações cadastrais do Mavericks. Por este motivo, possui um importante papel

de enriquecimento de dados de contato dos clientes. As informações do programa de fidelidade são relacionadas a categorias dos clientes no programa, resgates, saldo de pontos, pontos a expirar e etc.

3.1.2 Estudo e comparação entre modelos de dados

Tendo definido as fontes de informações a serem utilizadas, é necessário definir qual estratégia de modelagem das tabelas permite melhor desempenho. Como em diversos projetos de engenharia, existe um claro *trade-off* na escolha do modelo de dados. Não existe uma resposta pronta sobre qual melhor modelo de dados, portanto, é fundamental que se analise caso a caso. Esta é uma etapa de projeto fundamental em projetos de BI e não pode ser subestimada, pois erros nela significam um grande desperdício de recursos quando o projeto estiver em produção.

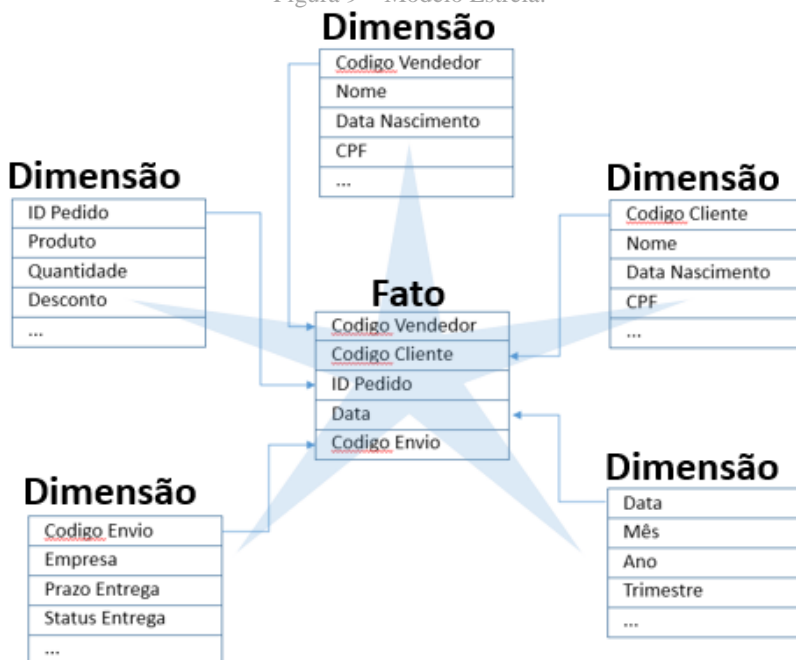
Por serem as abordagens mais aceitas na literatura e no mercado, foram analisadas as três seguintes modelagens de dados:

- Tabela Simples (*Flat Table*).
- Modelo Estrela (*Star Schema*).
- Floco de Neve (*Snowflake*).

Como sugere o nome, na modelagem do tipo tabela simples, todos os dados são armazenados em uma única tabela. Não se têm tabelas fatos e dimensões, são trazidos tanto dados transacionais como descritivos na mesma tabela. Por este motivo, possui um alto grau de desnormalização (redundância) de dados [17].

O modelo estrela consiste, basicamente, em organizar os dados de formas a se ter uma tabela com as principais informações (tabela fato), e outras tabelas complementares (tabelas de dimensões) conectando-se a tabela fato através de campos em comum (chaves de ligação). As tabelas de dimensões carregam as informações detalhadas sobre campos da tabela fato. Por este motivo, este modelo possui tabela fato normalizada (sem redundância). Quando visto de maneira gráfica, tem formato que remete a uma estrela, razão pela qual o modelo leva este nome. A figura 9 ilustra o modelo estrela.

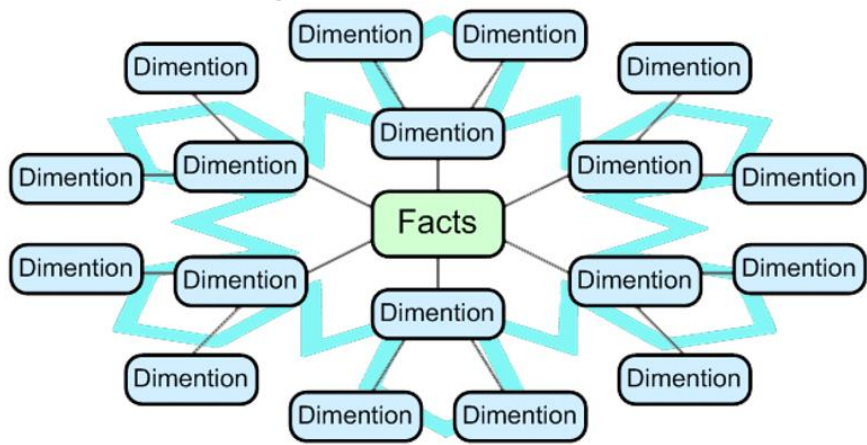
Figura 9 – Modelo Estrela.



Fonte: Elaborado pelo autor (2018).

O modelo floco de neve surge através da decomposição hierárquica (normalização) das tabelas de dimensões presentes no modelo estrela. Neste modelo, além das tabelas de dimensões ligadas diretamente à tabela fato, existem tabelas de dimensões ligadas às outras tabelas de dimensões. Conforme mostra figura 10, a análise gráfica deste modelo remete ao formato de um floco de neve, razão pela qual o modelo leva este nome.

Figura 10 – Modelo Floco de Neve.



Fonte: Adaptado a partir de [18].

Cada um dos modelos apresentados possui vantagens e desvantagens em relação aos outros modelos. Desta forma, é necessário que se escolha aquele que maximize as características buscadas ao mesmo tempo em que atenda aos requisitos do projeto. A figura 11 mostra uma comparação dos modelos sob cinco diferentes óticas, considerando um mesmo volume de informações.

Figura 11 – Comparação entre modelos de dados.

	Opção 1 Snowflake	Opção 2 Modelo Estrela	Opção 3 Tabela Simples
Tempo de Resposta			
Consumo de RAM			
Tempo execução Script			
Flexibilidade do Modelo			
Complexidade Script			

Fonte: Retirado de [19].

Sob os aspectos de tempo de resposta, destaca-se a performance da tabela simples. Isto ocorre devido a todas as informações estarem armazenadas em um único lugar, qualquer informação que precise ser acessada está disponível na memória RAM e não é necessário correlacionar tabelas para obter uma informação. O modelo estrela possui tempo de resposta levemente inferior porque algumas informações necessitam de uma chave de ligação para serem acessadas. Seguindo este raciocínio, o modelo floco de neve possui o pior dos tempos de resposta.

Já sob a ótica de consumo de memória RAM, o modelo mais otimizado é o floco de neve. Isto ocorre porque todas as informações estão normalizadas (não são redundantes) em todas as tabelas. Desta forma, não se carrega várias vezes a mesma informação para a memória RAM do servidor.

Em termos de tempo de execução do script de carga, o modelo estrela possui melhor desempenho. Ele destaca-se em relação à tabela simples, pois, sendo seminormalizado, trabalha com um volume muito menor de informações. Possui tempo de execução superior ao floco de neve, já que a interpretação de chaves de ligação entre tabelas de dimensão é complexa e requer alto esforço computacional.

Tanto o modelo estrela como tabela simples possuem ótimos índices de flexibilidade de modelo. Ou seja, suportam que o modelo de dados seja escalado a qualquer nível, desde que o servidor possua memória RAM suficiente.




Em relação ao nível de complexidade da programação do script, a tabela simples apresenta significativa superioridade em relação aos outros modelos. Isto se dá porque a tabela simples não possui chaves de ligação e sua implementação tende a ser simplificada.

3.1.3 Modelo de dados adotado

Tendo mapeado as fontes de dados a serem utilizadas e levando em consideração os *trade-offs* apresentados, foi montada uma matriz de decisão binária para escolha do modelo de dados ideal para o projeto. O objetivo de uma matriz de decisão binária é entender quais das alternativas são viáveis, sem fazer distinção qualitativa entre as opções. Sob cada critério, é analisado se a solução atende ou não atende aos requisitos do projeto. Caso mais de uma alternativa seja bem sucedida na análise binária de todos os critérios, é realizada uma análise qualitativa entre as possíveis alternativas.

A matriz de decisão é apresentada na figura 12.

Figura 12 – Matriz de decisão binária do modelo de dados a ser adotado

			
Consumo de RAM	✓	✓	✗
Complexidade do Script	✓	✓	✓
Tempo de Resposta	✗	✓	✓
Flexibilidade do Modelo	✗	✓	✓
Tempo de Execução do Script	✓	✓	✗

Fonte: Elaborado pelo autor (2018).

A partir da matriz de decisão, o modelo de dados escolhido e implementado foi o modelo estrela.

A tabela simples estouraria o limite de memória RAM do servidor (100 Gb), bem como levaria tempo demais para execução do script. Já o floco de neve, demoraria tempo demais para acessar indicadores com dimensões armazenadas longe da tabela fato e não se mostrou suficientemente escalável.

3.1.4 Extração dos dados

A primeira das etapas da ETL é a extração (*extract*). O objetivo desta etapa é extrair os dados brutos das bases e gravá-los em QVDs.

A principal vantagem de espelhar as informações de suas fontes primárias para QVDs é o alto ganho de desempenho das etapas posteriores e do banco de dados. QVD é um formato de arquivo no qual o QlikView processa informações com performance otimizada.

Abaixo, métodos de integração utilizados para cada uma das fontes de dados (métodos descritos em 3.2):

- Mavericks: Os dados deste ERP são armazenados em banco MYSQL. Trata-se de um banco de dados robusto, tradicional e com bom custo benefício. A extração de suas tabelas e respectivos dados foi realizada através de conexão ODBC (explicada em 2.3.1).
- Rabat e Padang: Integração por FTP (explicada em 2.3.2). Para configuração dos parâmetros de conexão, foi utilizado o software de conexão a FTP chamado FileZilla.

Para execução desta etapa, foi criada uma aplicação de QlikView exclusiva para extração dos dados. Como não possui objetos gráficos, a aplicação foi programada inteiramente por script. A figura 13 mostra trecho do script de extração.

Figura 13 – Trecho do código da extração do Mavericks (dados sigilosos foram ocultados).

```
//OLEDB CONNECT TO [Provide=
OLEDB CONNECT TO [Provider=

PRODUTO_TMP:
load
    codigoproduto as %key_cod_produto,
    codigoproduto as %key_cod_produto_novo,
    fornecedor1,
    referencial,
    descricao_basica,
    nome_setor,
    nome_linha,
    nome_tipologia,
    if (, nome_classificacao_produto) as nome_classificacao_produto,
    nome_formato,
    Fase_da_Vida,
    Tipologia,
    Formato_Real,
    unidade,
    portal as portal_produto
;
SQL
SELECT
    codigoproduto,
    fornecedor1,
    referencial,
    descricao_basica,
    nome_setor,
    nome_linha,
    nome_tipologia,
    nome_classificacao_produto,
    nome_formato,
    Fase_da_Vida,
    Tipologia,
    Formato_Real,
    portal,
    unidade
FROM
    export_produtos_tmp
where fornecedor1=7 and codigoproduto<1000000
;
NoConcatenate
PRODUTO:
LOAD
    minstring(%key_cod_produto) as %key_cod_produto,
    minstring(%key_cod_produto_novo) as %key_cod_produto_novo,
```

Fonte: Elaborado pelo autor (2018).

3.1.5 Transformação dos dados

A segunda e mais importante etapa da ETL é a transformação (*transform*). A aplicação de QlikView criada nesta etapa tem como entrada os QVDs gerados na etapa 3.1.4. Tem por finalidade gerar QVDs dos dados tratados (transformados), que servem para carga da aplicação final.

Esta aplicação é responsável pelos seguintes processos:

- Ajustar os dados brutos extraídos para a modelagem de dados escolhida, no caso, o modelo estrela.

Isto é realizado através do uso de funções de script do QlikView. Estas funções são da mesma família da linguagem SQL, apresentada em 2.2.1.

- Renomear campos das tabelas para nomes mais intuitivos.

Muitas vezes as sintaxes das tabelas de bancos de dados não são interpretáveis sem auxílio de *datasheet*. Desta forma, para facilitar a programação é necessário renomear as colunas importadas. Como boa prática, os nomes originais são substituídos por nomes compostos por prefixo e complemento. O prefixo deve ser abreviação da tabela de origem do dado e o complemento deve ser abreviação do nome da coluna. Por exemplo, a coluna de nome do produto da tabela Produtos foi renomeada de “XBLNR” para “Prod_Nome”.

- Criar campos calculados.

Campos calculados são colunas que possuem seus valores calculados a partir de valores de outras colunas existentes no modelo de dados.

Por exemplo, na base de dados de pedidos existe uma coluna referente à data dos pedidos. Porém, não existe uma coluna referente à data da última compra. Desta forma, nesta etapa é criada uma coluna nova que é preenchida de acordo com os valores encontrados na coluna de data dos pedidos. De acordo com o valor da

coluna de data da última compra, é calculada outra coluna, que é referente ao status de compra do cliente.

Figura 14 – Trecho do código da aplicação de transformação dos dados.

```
PROD:
NoConcatenate
Load
*,
If(WildMatch(Upper(DES_DEPARTAMENTO & DES_SEGMENTS & DES_SUB_SEGMENTS & DES_GRUPO
& DES_SUB_GRUPO), '*ACESSORIO*'), 'Acessório',
If(WildMatch(Upper(DES_DEPARTAMENTO & DES_SEGMENTS & DES_SUB_SEGMENTS &
DES_GRUPO & DES_SUB_GRUPO), '*CHINELO*'), 'Chinelo', 'Calçado')) as 1
IND_STATUS
Resident PROD3;

Drop Table PROD3;

Concatenate(PROD) 2
Load Distinct
'AREA.' & Right(COD_DEPARTAMENTO,1) as COD_CHAVE_PROD, 3
DES_DEPARTAMENTO
FROM [$(vGlobalDirDadoTransformado)/Dim_PROD_Mercadologica.qvd] (qvd);
```

Fonte: Elaborado pelo autor (2018).

O trecho de código da figura 14 destaca os processos listados de acordo com os índices da lista abaixo:

1. Criação de campo calculado.
2. Ajuste de dados para modelo estrela. Neste caso, realizado através da junção de duas tabelas de dimensões (desnormalização).
3. Renomeação de coluna.

3.1.6 Carga dos dados

A etapa de carga dos dados serve para importar para a aplicação de QlikView, na qual os indicadores serão gerados. São carregados todos os QVDs gerados na etapa anterior.

Esta etapa poderia ser executada na mesma aplicação na qual foi executada a transformação dos dados. Porém, como a transformação dos dados requer processamento para criação e alteração de dados, é recomendado que se separe a carga da transformação.

Além disso, como a aplicação final possui usuários que utilizarão as informações, é importante que os campos sejam renomeados de formas a tornar o trabalho com os dados mais intuitivos. O campo citado em 3.1.5 foi renomeado para “Nome do Produto”.

A figura 15 demonstra trecho do código da aplicação de carga de dados.

Figura 15 – Trecho do código da aplicação de carga dos dados.

```

1 $(must_include=lib://CNNDIR Includes/Parametro.inc);
2 $(must_include=lib://CNNDIR Includes/Transformação/Dimensão/T_Dim_Auto_Calendar.inc);
3
4 Dim_Centro:
5 LOAD
6     LOJA as [Loja - Código],
7     DES_UF as [Estado - Descrição],
8     GERENCIA as [Gerência - Descrição],
9     DIRETORIA as [Diretoria - Descrição],
10    "CSC - CML - Varejo" as [Estrutura Hierárquica],
11    AREA as [Área],
12    GRUPO as [Grupo],
13    AGRUP_NOM_CENTRO as [Agrupamento Centros],
14    COD_CENTRO as [Centro Lucro/Custo - Código],
15    DES_LOCAL as [Loja - Local],
16    LOJA&'-'&Capitalize(DES_LOCAL) as [Loja - Descrição],
17    DES_EMPRESA as [Empresa],
18    DES_CIDADE as [Loja - Cidade],
19    %key_mapa,
20    "High/Low"
21 FROM [lib://CNNDIR Dados/Transformados\Dim_Centro_Farol_Lojas.qvd]
22 (qvd);

```

Fonte: Elaborado pelo autor (2018).

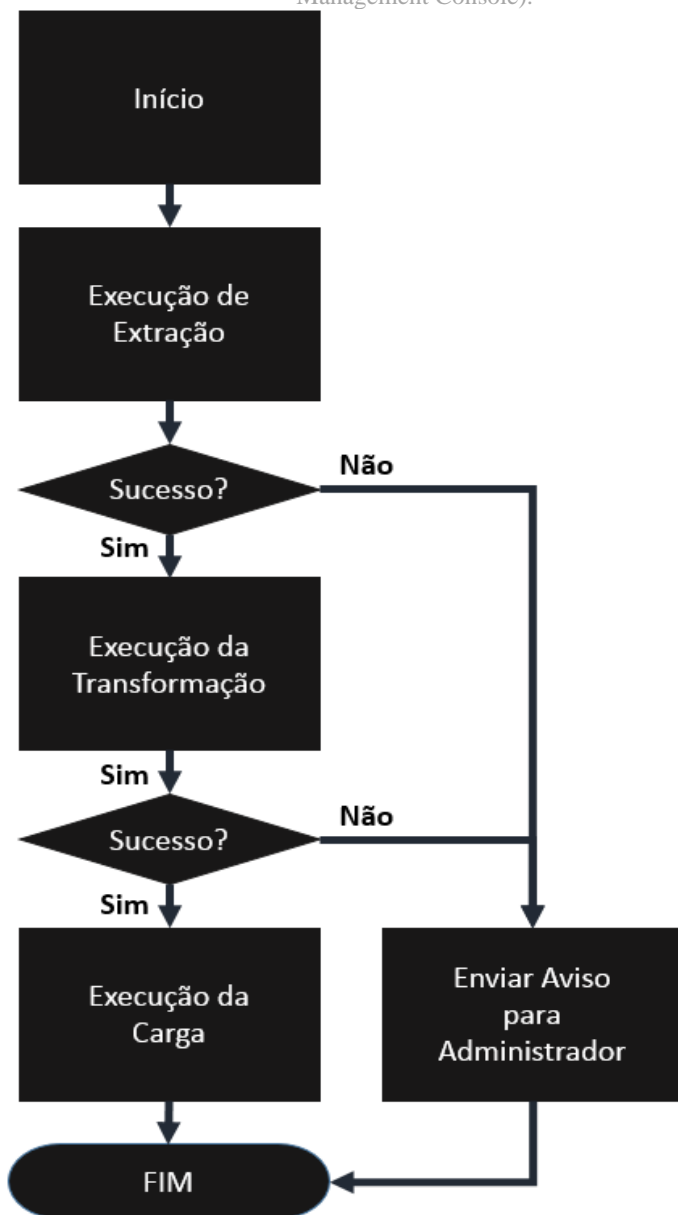
3.1.7 Automação do processo de ETL

Como o projeto visa tomada de decisões, é necessário que as informações estejam sempre atualizadas diariamente. Para que não haja dependência de pessoas e sim de processos, as etapas descritas em 3.1.4 a 3.1.6 foram automatizadas.

Para tal, o QlikView possui módulo de gerenciamento de aplicações (QMC). Neste módulo, são configuradas através de interface gráfica as condições de contorno necessárias para que as tarefas sejam executadas.

O fluxograma de atualização dos dados implementado é apresentado na figura 16. O processo foi configurado para iniciar diariamente às 04h00min.

Figura 16 – Fluxograma de atualização dos dados configurado no QMC (Qlik Management Console).



Fonte: Elaborado pelo autor (2018).

3.2 Modelagem de negócio

Tendo definido modelagem de dados do projeto, foi iniciada etapa de modelagem de negócio. O objetivo desta etapa é especificar os entregáveis do projeto, é definir o que de fato chegará como entrega palpável ao usuário final.

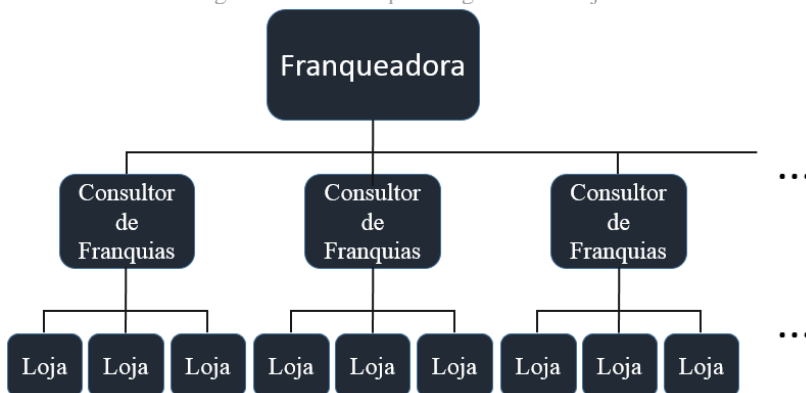
É de extrema importância que esta etapa seja feita em conjunto com os gestores e analistas da área, pois requer conhecimentos específicos do modelo de negócio. É nesta etapa que a inteligência humana é usada para transformar dados em *outputs* valiosos para o negócio. Se esta etapa não for bem feita, de nada adianta para o sucesso do projeto uma base de dados bem modelada.

3.2.1 Levantamento de informações e entrevistas

A primeira das etapas da modelagem de negócio é levantar o máximo de *inputs* (informações) sobre a área na qual o projeto vai ser executado. O papel do entrevistador é entender as dificuldades citadas, mapear dificuldades que existem, mas não são percebidas, e as oportunidades existentes.

Nesta etapa deve ser ouvido o máximo de pessoas que possuem interface com o negócio em questão. Desta forma, é possível garantir que os fatos foram analisados sob todas as óticas do negócio. A figura 17 ajuda a entender como é o organograma da gestão das lojas.

Figura 17 – Hierarquia de gestão das lojas.



Fonte: Elaborado pelo autor (2018).

Pensando nisso, foram entrevistadas 19 pessoas que atuam em todos os níveis do negócio. Entre elas:

- Gerente, coordenador e analistas da franqueadora: Equipe responsável pelas áreas na qual o projeto foi executado.
- Consultores de franquias: São profissionais contratados pela franqueadora. Para cada grupo de lojas (em média 15) existe um consultor de franquias. São responsáveis por garantir que exista um padrão entre as lojas da rede. Por fazerem o intermédio entre loja e franqueadora, conhecem a fundo tanto a operação da loja como da franqueadora. Sendo assim, possuem um importante papel na etapa de levantamento de informações. Oito consultores de franquias foram entrevistados.
- Franqueados, gerentes e vendedores: São as pessoas que estão na operação diária das lojas, na ponta do processo. Possuem um importante papel no processo de levantamento de informações. Para que o projeto tenha aceitação em massa, a solução implementada precisa ser aderente às necessidades e oportunidades mapeadas com estas pessoas.

3.2.2 Interpretação das entrevistas

O intuito desta etapa é traduzir todas as informações levantadas em problemas palpáveis de serem solucionados.

Ao final da etapa de entrevistas e levantamento de informações, foram compiladas todas as respostas. Após análise, foi montado um documento listando as principais dificuldades e oportunidades encontradas. Por questões de sigilo de projeto, este documento não será disponibilizado.

Esta etapa é importante, pois garante definição dos problemas que precisam ser resolvidos. Define um rumo para o projeto. Durante as próximas etapas do desenvolvimento, sempre se procurou tomar decisões que fossem aderentes aos *outputs* dessa etapa.

3.2.3 Definição dos indicadores (KPIs)

KPI é o acrônimo para *Key Performance Indicator*. Em tradução livre para o português, significa Indicador Chave de Desempenho.

Indicadores de desempenho são métricas quantitativas que refletem o desempenho de uma organização na realização de seus objetivos e estratégias [20].

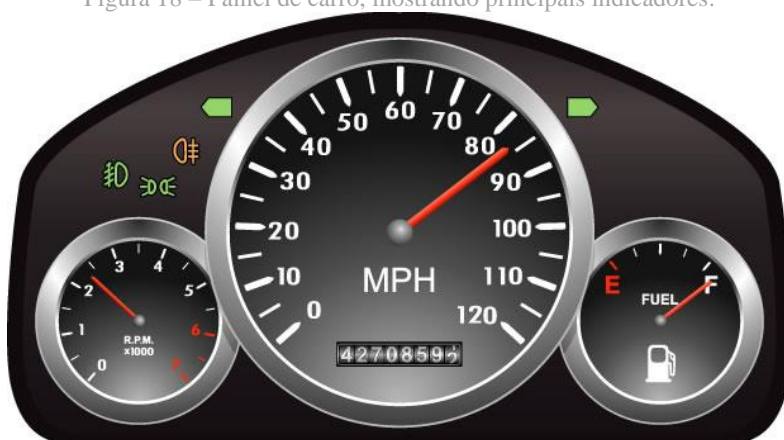
São funções matemáticas aplicadas a um conjunto de dados, no qual o resultado das operações revela uma importante métrica sobre o sistema analisado. Por definição, não precisam seguir algum padrão e podem ser estabelecidos de forma a auxiliar no processo que se deseja atuar.

Esta flexibilidade de uso de diferentes nomenclaturas e adaptações a indicadores resulta em um cenário com um vasto número de possíveis indicadores. Este grande volume de indicadores tornaria inviável a análise dos mesmos e acabaria atrapalhando a tomada eficiente de decisão. Por este motivo, é necessário escolher quais indicadores são chaves para demonstrar o desempenho da empresa [20].

A importância de mostrar somente os principais indicadores a serem analisados fica clara quando se faz analogia com um carro. Em um carro existem inúmeros sensores e medições sendo realizadas a cada instante. Desta forma, seria possível gerar centenas de indicadores sobre um carro.

No entanto, como é ilustrado na figura 18, os projetistas de painéis de carros mostram somente os indicadores mais relevantes para tomada de decisão do motorista. No caso da figura, o carro mostra os indicadores: RPM do motor, velocidade, odômetro e quantidade de combustível no tanque.

Figura 18 – Painel de carro, mostrando principais indicadores.



Fonte: Retirado de [21].

Sendo assim, esta etapa foi focada em definir os principais indicadores do projeto. Todos os indicadores foram pensados de forma a auxiliarem na tomada de decisão e operação de vendas das lojas, visando crescimento em vendas (*VNPE*) e carteira ativa (*CAE*). Foram selecionados 19 indicadores, que não são explicados individualmente por questões de sigilo de projeto. No entanto, os indicadores podem ser divididos nas seguintes categorias:

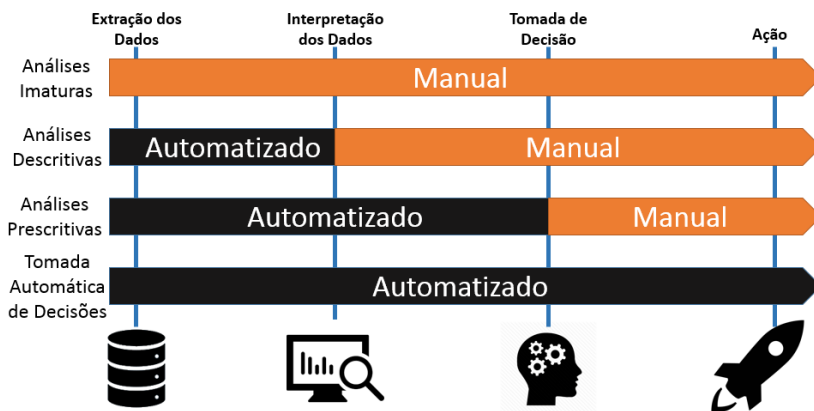
- Vendas;
- Relacionamento;
- Potencial para novos projetos;
- Programa de relacionamento.

3.2.4 Análises prescritiva

Como as lojas lidam com uma grande carteira de clientes (*CC*), chegando a cinco mil por loja, é essencial que os seus gestores consigam priorizar quais ações com quais clientes trarão mais resultados. Porém, em um universo com tantos clientes e indicadores fica difícil de priorizar as oportunidades rapidamente.

A figura 19 demonstra as etapas entre dados brutos para diferentes níveis de maturidade de automação em *data analytics* (ciência da análise de dados).

Figura 19 – Níveis de maturidade em data analytics e respectivos processos manuais e automatizados.



Fonte: Elaborado pelo autor (2018).

Pensando na necessidade de tomada rápida de decisão dos gestores das lojas, foi definido que o projeto não poderia ser apenas de análises descritivas. Ou seja, não poderia somente levar as informações até os gestores. Deveria também ter interpretação automatizada dos dados, facilitando a tomada de decisão.

Para levar o projeto a este nível de maturidade das análises, foram criadas duas funcionalidades nos relatórios desenvolvidos:

- Alertas dos KPIs.
- Menu de ações.

3.2.4.1 Alertas dos KPIs

Os alertas foram criados com intuito de ser uma forma fácil de mostrar principais *gaps* (lacunas) e oportunidades para os gestores das lojas.

Seu funcionamento é muito simples. São funções que foram programadas para que quando os indicadores tenham determinado comportamento, seja criado um alerta para a loja sobre aquela oportunidade.

Para exemplificar, suponhamos que determinado cliente possua pontos a expirar no programa de relacionamento das lojas. Caso não houvesse um alerta de oportunidade, o próprio gestor teria que analisar as informações e perceber que este cliente possui pontos a expirar. Com

a criação dos alertas, foi definido que junto com as demais informações deste cliente, a loja receberia um alerta que o mesmo possui pontos a expirar. A partir deste ponto, o vendedor deveria entrar em contato com o cliente e executar as abordagens comerciais para este tipo de oportunidade.

3.2.4.2 Menu de ações

Como as lojas lidam com um grande volume de clientes, é necessário um nível de abstração superior de informações. Os gestores da loja precisam ter uma visão macro da situação desta. Assim, ficam aptos a atuarem de maneira guiada às principais oportunidades da loja, melhorando a captura de oportunidades ao nível de vendedores e clientes.

Para fornecer esta visão macro das oportunidades da loja, foi criado o menu de ações. Este é gerado a partir de uma análise automática dos indicadores da loja. Por conter informações confidenciais e estratégicas das diretrizes de ação da empresa, o menu de ações não pôde ser demonstrado em figuras.

3.3 Implantação prática do modelo

Esta seção apresenta os desenvolvimentos realizados para implementação do modelo definido na etapa 3.2.

Os desenvolvimentos foram feitos com base no QlikView e Nprinting, ferramentas apresentadas no capítulo 2.

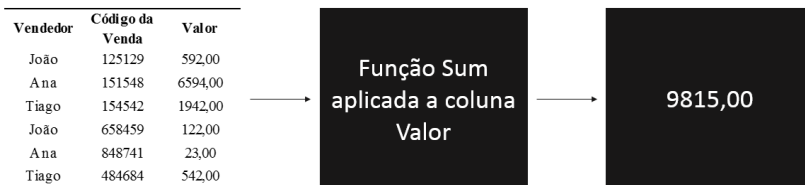
3.3.1 Programação dos indicadores

Com a utilização de funções pré-definidas do QlikView, são criados indicadores.

Um exemplo de função pré-definida do QlikView é a função *Sum* (soma). Esta função soma todos os valores passados como parâmetros e retorna o resultado dos respectivos valores encontrados.

A figura 20 ilustra o funcionamento desta função pré-definida. Define-se qual a coluna de entrada, e a função retorna como saída o resultado da soma de todos os valores encontrados. Porém, em geral, as regras de indicadores são mais complicadas que somas simples.

Figura 20 – Exemplo da função Sum, disponível para uso em QlikView.



Fonte: Elaborado pelo autor (2018).

Para tal, o QlikView permite que seja realizada combinação de funções. Com a combinação de funções, que individualmente são simples, são desenvolvidos indicadores com regras complexas.

A figura 21 mostra um exemplo com utilização de diversas funções combinadas. O código da figura representa um indicador de previsão de vendas e foi desenvolvido neste projeto.

Figura 21 – Exemplo funções combinadas para gerar indicador de previsão de vendas.

```
$(VKPI_VNP_ESP_MES_ATUAL)+
if(month(max({<nome_consultor=>,nome_vendedor=>,referencial=>,descricao_basica=>,empresa_nome_fantasia=>,
nome_tipologia=>,nome_setor=>})data_documento))-month(reloadtime())) and
year(max({<nome_consultor=>,nome_vendedor=>,referencial=>,descricao_basica=>,empresa_nome_fantasia=>,
nome_tipologia=>,nome_setor=>})data_documento))-year(reloadtime())) ,
if(max(data_documento)>monthstart(now(0)),
(
// RT sem SER
SUM({<cod_especificador=>{''},
Ano=>,Mês=>,Dia=>,DiasFaltantes=>{'<$(=MAX({<nome_classificacao_produto=>, nome_setor=>, referencial=>,
nome_vendedor=>, nome_especificador=>, nome_franqueado=>, empresa_nome_fantasia=>, nome_consultor=>,
descricao_basica=>, unidade=>,Dia=>,Mês=>,Ano=>,data_documento=>{'<$(=date(now(0)-1,'DD/MM/YYYY'))'})>
DiasFaltantes'))'},data_documento=>{'>=$(=monthstart(addmonths(date(now(0),'DD/MM/YYYY'),-3)))<$(=
monthend(addmonths(date(now(0),'DD/MM/YYYY'),-1)))'},cancelado='N'}, [Tipo de Registro]='Pedido'
})
> } if( isnull(documento_arquiteto_ser), preco_unitario * qtdeproduto -desconto_proporcional,0) )
//SEM sem RT
=
SUM({<documento_arquiteto_ser=>{''},
Ano=>,Mês=>,Dia=>,DiasFaltantes=>{'<$(=MAX({<nome_classificacao_produto=>, nome_setor=>, referencial=>,
nome_vendedor=>, nome_especificador=>, nome_franqueado=>, empresa_nome_fantasia=>, nome_consultor=>,
descricao_basica=>, unidade=>,Dia=>,Mês=>,Ano=>,data_documento=>{'<$(=date(now(0)-1,'DD/MM/YYYY'))'})>
DiasFaltantes'))'},data_documento=>{'>=$(=monthstart(addmonths(date(now(0),'DD/MM/YYYY'),-3)))<$(=
monthend(addmonths(date(now(0),'DD/MM/YYYY'),-1)))'},cancelado='N'}, [Tipo de Registro]='Pedido'
})
> } if( isnull(cod_especificador), preco_unitario * qtdeproduto -desconto_proporcional,0) )
//ambos
=SUM({<cod_especificador=>{''},
Ano=>,Mês=>,Dia=>,DiasFaltantes=>{'<$(=MAX({<nome_classificacao_produto=>, nome_setor=>, referencial=>,
nome_vendedor=>, nome_especificador=>, nome_franqueado=>, empresa_nome_fantasia=>, nome_consultor=>,
descricao_basica=>, unidade=>,Dia=>,Mês=>,Ano=>,data_documento=>{'<$(=date(now(0)-1,'DD/MM/YYYY'))'})>
DiasFaltantes'))'},data_documento=>{'>=$(=monthstart(addmonths(date(now(0),'DD/MM/YYYY'),-3)))<$(=
monthend(addmonths(date(now(0),'DD/MM/YYYY'),-1)))'},cancelado='N'}, [Tipo de Registro]='Pedido'
})
> } if(not isnull(documento_arquiteto_ser), preco_unitario * qtdeproduto -desconto_proporcional,0) )
/ 3),0)
```

Fonte: Elaborado pelo autor (2018).

Seguindo processo de combinação de funções foram desenvolvidos 19 indicadores, definidos na etapa 3.2.3.

3.3.2 Programação do Nprinting

Esta subseção apresenta os tópicos referentes ao uso do Nprinting no projeto, ferramenta utilizada para distribuição automatizada dos relatórios.

No caso deste projeto, as principais razões para enviar as informações para os gestores ao invés de exibí-las no QlikView são:

- Eficiência no consumo da informação.
Para o usuário, receber um e-mail é um processo passivo. Acessar um sistema para visualizar relatórios é um processo ativo. Desta forma, retirando a necessidade de acesso ao QlikView e entregando relatórios semanalmente por e-mail, é estimulado o consumo da informação.
- Economia financeira.
Como os relatórios precisam ser disponibilizados para mais de 151 gestores de lojas, caso a informação não fosse enviada por e-mail cada loja precisaria de pelo menos uma licença de acesso ao QlikView. Cada licença custa cerca de R\$ 8 mil. O Nprinting custa cerca de R\$ 70 mil e suporta distribuição para todos os usuários finais. Desta forma, a implementação da distribuição via e-mail permite economia financeira de mais de R\$ 1 milhão*.

A figura 22 demonstra estratégia de distribuição automatizada das análises criadas através do Nprinting. As etapas 2 e 3 são apresentadas respectivamente em 3.3.2.1 e 3.3.2.2.

Figura 22 – Estratégia de distribuição automatizada de relatórios através do uso do Nprinting.



Fonte: Elaborado pelo autor (2018).

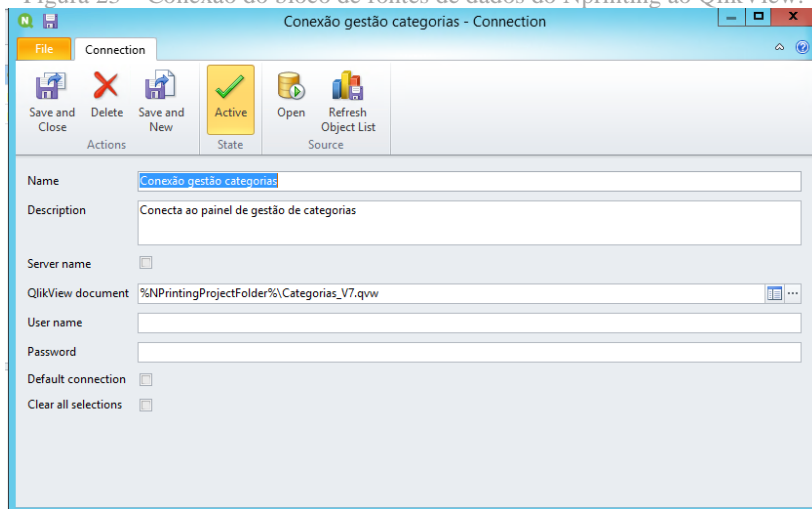
* Sem considerar possíveis descontos nos preços originais das ferramentas

3.3.2.1 Criação dos relatórios

O processo de criação dos modelos de relatórios no Nprinting (etapa 2 da figura 22) é realizado através do uso dos blocos de fontes e de relatórios.

O bloco de fontes realiza a conexão dos objetos gerados no QlikView com o Nprinting. Permite conexão a uma ou mais aplicações do QlikView, o que lhe garante eficiência na reutilização de códigos. A conexão à aplicação desenvolvida para este projeto é apresentada na figura 23.

Figura 23 – Conexão do bloco de fontes de dados do Nprinting ao QlikView.



Fonte: Elaborado pelo autor (2018).

Após carga dos objetos para o Nprinting, são posicionados os objetos de maneira desejada de acordo com o tipo de relatório. Neste caso, foi criado um relatório em Excel.

A figura 24 mostra trecho do posicionamento dos objetos importados em relatório de Excel. Quando o modelo gera um relatório automatizado, as informações definidas como variáveis são substituídas por seus respectivos valores, definidos no QlikView.

Figura 24 – Criação de relatório em Excel através do uso do bloco de relatórios do Nprinting.

<vLOJA_1>

Lista de Especificadores 'Pronta para Ação' - <vDataReload>

Para selecionar os Especificadores a ser As oportunidades de abordagem para c

Tipo de Ação		Identificação		
<Tipo de Ação>	<CPF>	<Posição no Ranking da Loja (RS)>	<Vendedor Responsável>	<Especificador>
<Tipo de Ação>	<CPF>	<Posição no Rai>	<Vendedor Responsável>	<Especificador>

Fonte: Elaborado pelo autor (2018).

3.3.2.2 Distribuição automatizada dos relatórios

Figura 25 – Configuração do e-mail de envio dos relatórios.

FileReport Task

Save and Close

Delete

Save and New

Active

General

Reports

Recipients

Filters

Message

Name

Lista Pronta Especificadores

Description

Connections

Email

Send Email

✓

From

Full N

Add CC/BCC

✓

CC

BCC

Add/Replace

Replace options default e-mail addresses with these e-mail addresses (CC and BCC)

Send reports to these e-mail addresses in addition to options default e-mail addresses (CC and BCC)

Subject

Lista de Especificadores %Recipient_Variable_Conexão gestão categorias_vLOJA/%

Attachments

✓

Reports/Files

Lista de Especificadores - Pronta para Ação (Excel Workbook)

Format

Text

HTML Text

HTML File

Body

Caro Franqueado(a) %Recipient_Variable_Conexão gestão categorias_vNOM_FRANQUEADO%

As listas de Especificadores 'Prontas para a Ação' estão sendo muito bem recebidas pela Rede Portobello Shop e tem apoiado as lojas a aprimorarem seus planos de ação.

Confira anexo sua lista de Especificadores 'Pronta para Ação' do mês de %Variable_Conexão gestão categorias_vMesAtual% e aproveite as oportunidades da sua carteira.

Teremos prazer em receber suas dúvidas, críticas ou sugestões sobre as 'Listas Prontas para Ação' através do email %Recipient_Variable_Email%

Ótimas vendas para Especificadores!

Fonte: Elaborado pelo autor (2018).

Para realizar distribuição automatizada das informações é necessário configurar os parâmetros de envio das informações. A figura 25 demonstra a configuração dos parâmetros do e-mail a ser enviado. Nesta tela, são configurados todos os pontos referentes aos disparos, desde recipientes até quais relatórios devem ser anexados ao e-mail.

3.4 Validação e ajustes

Ao final das etapas de desenvolvimento, finalizadas em 3.3, foi realizada validação do desenvolvimento com os envolvidos na modelagem de negócio.

O objetivo desta etapa é validar se o projeto desenvolvido:

- É aderente ao modelo especificado.
- Atinge os objetivos gerais do projeto.

Para tal, foi apresentado protótipo das entregas realizadas para os gestores das áreas envolvidas da empresa. Os *feedbacks* coletados foram implantados na ferramenta.

3.5 Lançamento

Para garantir que quando a ferramenta fosse lançada para as 151 lojas da rede a mesma fosse isenta de erros, o lançamento do projeto foi realizado em duas etapas.

Na primeira, a ferramenta foi lançada em piloto com uma amostragem reduzida de lojas. Vinte lojas foram selecionadas para, por um período de um mês, terem acesso à ferramenta em primeira mão.

Em contrapartida, estas lojas se comprometeram a usar exaustivamente a plataforma neste período e relatar quaisquer desconfortos encontrados. Nesta etapa, por não estarem hospedado nos servidores de produção da empresa, alguns processos ainda precisavam ser executados manualmente.

Passado o piloto de um mês sem grandes ajustes necessários, o projeto foi implantado nos servidores de produção da empresa. A partir desta implantação, todos os processos passaram a ser executados de maneira automatizada e sem intervenção humana.

4 RESULTADOS

Nesta seção, são apresentados os resultados obtidos com a implementação prática do projeto. A seção apresenta os tópicos divididos em três frentes: questões de tempo e eficiência do projeto, alcance do projeto e resultados financeiros.

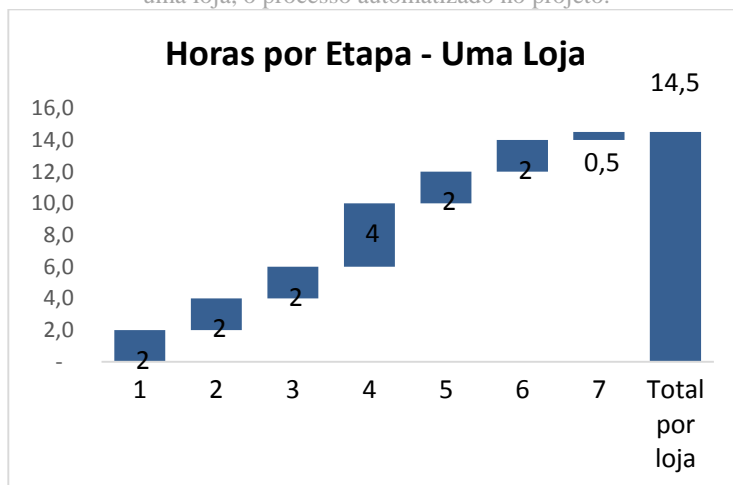
4.1 Eficiência e tempo de processamento

O projeto executado implementou um processo totalmente novo e sem precedentes na empresa. Portanto, não é possível afirmar com exatidão qual foi o ganho de tempo para execução dos processos como um todo. Nas subseções a seguir, é apresentada uma estimativa de quanto tempo seria necessário para executar o processo automatizado caso se desejasse executar o processo manualmente e, em seguida, realiza-se comparação com os resultados de tempo e eficiência obtido com o projeto automatizado.

4.1.1 Estimativa de tempo para execução manual

Estima-se que caso se executasse manualmente o processo automatizado, para uma loja, seriam necessárias cerca de 14,5 horas, conforme mostra a figura 26:

Figura 26 – Estimativa de tempo necessário para executar manualmente, para uma loja, o processo automatizado no projeto.



Fonte: Elaborado pelo autor (2018).

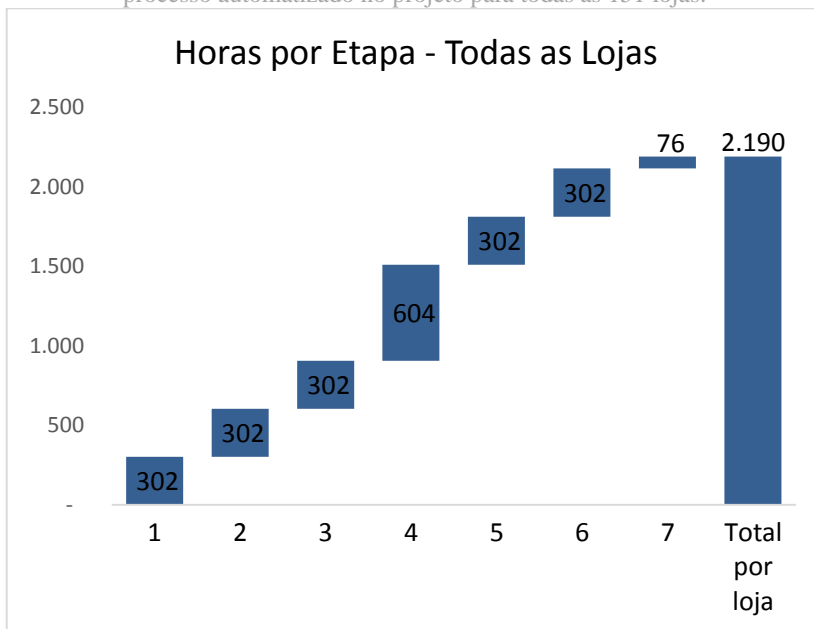
Para melhor visualização da figura, foram substituídos os nomes das etapas por números. Abaixo, o correspondente de cada um dos números das etapas das figuras 26, 27 e 28.

Etapas e respectivas descrições:

1. Aquisição dos dados
2. Modelagem dos dados
3. Integração entre fontes
4. Geração dos indicadores
5. Identificação das oportunidades
6. Geração do menu de ações
7. Distribuição por e-mail

Extrapolando-se o tempo necessário para execução do processo de uma loja para as 151 lojas da rede, seriam necessárias 2.190 horas (131.400 minutos) de trabalho para execução do processo, conforme mostra figura 27.

Figura 27 – Estimativa de tempo necessário para executar manualmente o processo automatizado no projeto para todas as 151 lojas.



Fonte: Elaborado pelo autor (2018).

Isto equivale a um total de 274 dias (39 semanas) de trabalho, considerando-se jornadas de trabalho de 8 horas por dia útil. Um ano possui, em média, 252 dias úteis. Ou seja, seria necessário mais de um ano de trabalho de uma pessoa dedicada exclusivamente a isto para executar este processo uma única vez.

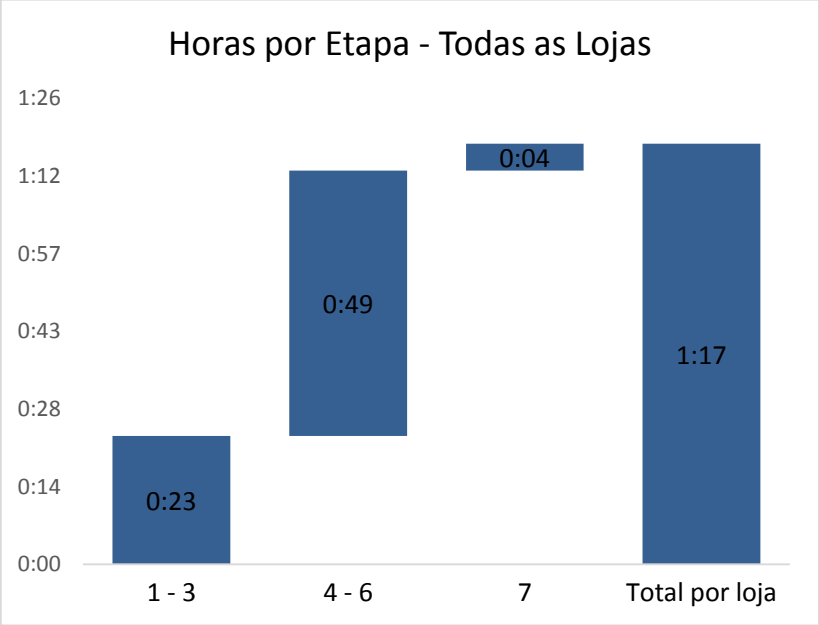
Dado que após o projeto ser implementado o processo passou a ser executado semanalmente, caso se executasse manualmente esta rotina seriam necessárias 39 pessoas executando o processo continuamente para distribuir as informações para as lojas. Além de ser inviável o número de pessoas que precisariam estar continuamente realizando este trabalho, as informações chegariam com atraso e com grande chance de conter erros humanos. Desta forma, fica claro que caso o projeto implantado não fosse automatizado, não seria factível executar o processo manualmente.

4.1.2 Análise de eficiência e tempo de execução do processo automatizado

Após a implantação do projeto, todos os processos implementados descritos no capítulo 3 e na subseção 4.1.1 deste trabalho foram completamente automatizados. Isto significa que com a implantação deste projeto, além de permitir a execução de processos que não seriam possíveis de se executar manualmente, foi criado um processo completamente independente de pessoas. Além de trazer economia financeira e de tempo, um processo independente de pessoas implica em um processo livre de erros humanos, o que seria praticamente inevitável ao se trabalhar com um volume tão grande de informações.

A figura 28 mostra o tempo total de execução do processo automatizado para as 151 lojas:

Figura 28 – Tempo de execução das etapas do processo automatizado para todas as 151 lojas.



Fonte: Elaborado pelo autor (2018).

Analisando a figura 28, observa-se que, para executar todas as atividades envolvidas no processo, foi necessário 1 hora e 17 minutos.

Comparando-se este resultado com as 2.190 horas que seriam necessárias para executar o processo manualmente, este resultado traz um ganho de eficiência de mais de 1.700 vezes, conforme mostra tabela 2.

Tabela 2 – Ganho de eficiência obtido com automatização do processo desenvolvido no projeto.

Execução Manual (em minutos)	Execução Automatizada (em minutos)	Ganho de Eficiência (%)
131400,00	77,00	170549%

Fonte: Elaborado pelo autor (2018).

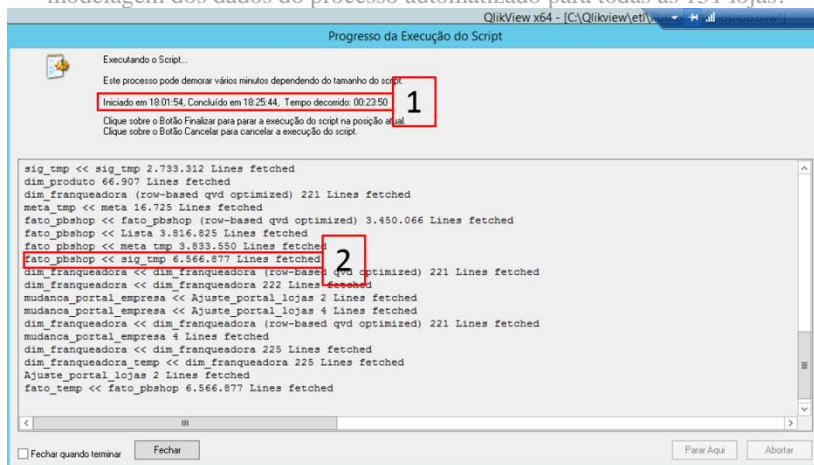
Conforme se pode ver na figura 28, os maiores tempos desprendidos no processo automatizado ocorrem nas etapas 1 a 3 e 4 a

6. Isto ocorre porque, como o volume de dados é muito grande, é necessário um alto esforço computacional para extração e processamento dos dados.

A figura 29 mostra o resultado do processo automatizado para as etapas 1 a 3, executado no QlikView. Trata-se do processo de extração, modelagem e integração de mais de 6,5 milhões de linhas e 130 colunas de dados. Nesta figura, destacam-se dois pontos, que representam respectivamente:

- Ponto 1: tempo total de execução das etapas, que foi de 23 minutos e 50 segundos.
- Ponto 2: número de linhas da tabela fato do modelo de dados. Esta é a maior tabela e ficou com total de 6.566.877 linhas em uma única tabela.

Figura 29 – Tempo de execução das etapas de extração, integração e modelagem dos dados do processo automatizado para todas as 151 lojas.



Fonte: Elaborado pelo autor (2018).

As figuras 30 e 31 mostram o resultado do processo automatizado das etapas 4 a 6, executado no Nprinting. Estas etapas contemplam processo de geração dos indicadores, oportunidades e menu de ações. As figuras mostram respectivamente início e fim do processo, ambos destacados em vermelho. Analisando as figuras, nota-se que o tempo total de execução para estas etapas foi de 49 minutos e 10 segundos.

Figura 30 – Início do processo automatizado de geração dos indicadores, dos alertas e do menu de ações.

```
//INÍCIO CRIAÇÃO RELATÓRIOS + OPORTUNIDADES + MENU DE AÇÕES
06/12/2016 13:15:30 Message Start Execute Task
06/12/2016 13:15:30 Message Apply filter: CASCAREL
06/12/2016 13:15:31 Message Filter field: empresa_nome_fantasia {CASCAREL} (
06/12/2016 13:15:34 Message Start Lista de Especificadores - Pronta para Ação
06/12/2016 13:15:34 Message Report type: Excel Workbook (Excel Workbook)
```

Fonte: Elaborado pelo autor (2018).

Figura 31 – Fim do processo automatizado de geração dos indicadores, dos alertas e do menu de ações.

```
\TAUBATE_000140-F00108_12 2016.xlsx)
06/12/2016 14:04:39 Message End Lista de Especificadores - Pronta para Ação
06/12/2016 14:04:40 Message End Execute Task
//FIM CRIAÇÃO RELATÓRIOS + OPORTUNIDADES + MENU DE AÇÕES
```

Fonte: Elaborado pelo autor (2018).

Por fim, resta a etapa 7, de distribuição automatizada por e-mail. Conforme mostram as figuras 32 e 33, esta etapa levou um total de 4 minutos e 59 segundos para ser executada. Foram enviados e-mails com os arquivos gerados nas etapas anteriores para 151 lojas, totalizando 453 destinatários inclusos.

Figura 32 – Início do processo automatizado de distribuição por e-mail dos relatórios.

```
//INÍCIO DISTRIBUIÇÃO POR E-MAIL
06/12/2016 14:04:40 Message Start Output Distribution
06/12/2016 14:04:40 Message Connecting to mail server using TLS (Transport Layer Security)
06/12/2016 14:04:41 Message Copying file: C:\Users\Public\Documents\NPrinting\Output\SP -
```

Fonte: Elaborado pelo autor (2018).

Figura 33 – Fim do processo automatizado de distribuição por e-mail dos relatórios.

```
06/12/2016 14:09:39 Message Message Sent!
06/12/2016 14:09:39 Message Disconnecting from mail.portobello.com.br
06/12/2016 14:09:39 Message Disconnected from mail.portobello.com.br
06/12/2016 14:09:39 Message End Output Distribution
//FIM DISTRIBUIÇÃO POR E-MAIL
```

Fonte: Elaborado pelo autor (2018).

4.2 Alcance do projeto

O projeto apresentado teve alcance nacional e foi reconhecido nacionalmente pela empresa na qual foi executado.

Entre os principais resultados de alcance do projeto, destaca-se a melhoria de vida e o alcance direto a:

- 151 lojas da rede de franquias;
- 121 franqueados;
- 900 funcionários, entre gerentes, vendedores e outros.

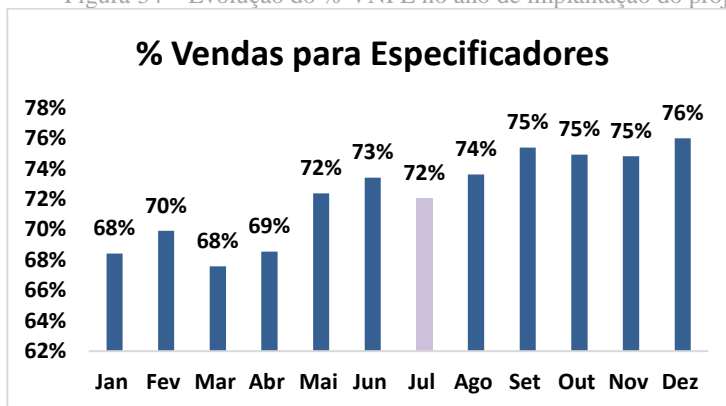
4.3 Resultados financeiros

Esta seção apresenta os resultados do projeto em termos financeiros. Estes resultados foram considerados como excelentes e acima da expectativa dos gestores das áreas envolvidas.

Sabe-se que os resultados não são exatos, pois o faturamento de uma rede de franquias com faturamento superior a R\$ 600 milhões e mais de 150 lojas depende de diversas variáveis. Porém, dado que este foi o principal projeto executado no período pelas principais áreas responsáveis pelo resultado financeiro da empresa, considera-se que grande parte do resultado global obtido pela empresa é proveniente deste projeto.

O projeto foi implantado e disponibilizado para as lojas da rede de franquias no final de Julho de 2016. Nos seguintes resultados apresentados, destaca-se o mês de implantação na cor cinza, a fim de facilitar visualização dos resultados obtidos a partir da data de implantação do projeto. A figura 34 mostra a evolução do indicador “% VNPE” (% de vendas para especificadores).

Figura 34 – Evolução do % VNPE no ano de implantação do projeto.



Fonte: Elaborado pelo autor (2018).

Analisando a figura 34, destaca-se:

- Nos meses anteriores à implantação do projeto (de Janeiro a Junho), a média do indicador % VNPE era de 70%. Logo no primeiro mês após implantação do projeto (Agosto), o indicador disparou e a empresa obteve índices nunca obtidos antes, fechando o mês em 74%.
- Em todos os meses após implantação do projeto, a empresa obteve resultados superiores a todos os resultados já obtidos nos períodos anteriores a implantação do projeto.
- No decorrer do ano, o indicador atingiu teto de 76% e fechou com média de 75% no período de Agosto a Dezembro.

O faturamento da empresa nos meses de Julho a Dezembro foi na ordem de 300 milhões de reais. A tabela 3 mostra a diferença estimada de faturamento obtida através da implementação do projeto.

Tabela 3 – Estimativa de diferença de VNPE obtida através da implementação do projeto.

Faturamento Total de Jul- Dez	% VNPE Pré projeto	VNPE Sem Projeto	% VNPE Pós Projeto	VNPE Com Projeto	Diferença
300 milhões	70%	210 milhões	75%	225 milhões	15 milhões

Fonte: Elaborado pelo autor (2018).

Como demonstra a tabela 3, entre Julho e Dezembro, estima-se que a empresa obteria um resultado de R\$ 210 milhões em vendas para especificadores caso o projeto não fosse implementado. Com a implementação do projeto e a ascensão do percentual de vendas para especificadores para 75%, a empresa obteve R\$ 225 milhões em vendas para especificadores no período.

Conforme mostra tabela 4, este ganho de R\$ 15 milhões em vendas para especificadores representa um crescimento de 7,14% no indicador de vendas para especificadores.

Tabela 4 – Crescimento do indicador de VNPE obtido através da implementação do projeto.

VNPE Sem Projeto	VNPE Com Projeto	Diferença	Crescimento
210 milhões	225 milhões	15 milhões	7,14%

Fonte: Elaborado pelo autor (2018).

5 CONCLUSÃO

Todo este projeto foi implementado em uma empresa do ramo de rede de franquias. Como qualquer outra empresa de grande porte e alto faturamento, esta possui desafios a serem solucionados e oportunidades para melhorias contínuas.

Neste âmbito, o projeto teve um importante papel, pois conseguiu por meio de técnicas de engenharia e gestão melhorar a eficiência da organização. Através dos resultados, ficou evidente o impacto causado na vida de centenas de pessoas.

Como última etapa antes da formação como engenheiro eletricista, o projeto foi importante, pois permitiu contato com problemas reais que necessitam de soluções reais.

Um ponto que felizmente foi aplicado e será levado como lição deste trabalho, é a importância de seguir uma metodologia de projetos e etapas bem definidas.

Vale destacar também a importância do trabalho em grupo. Sem os especialistas de negócios envolvidos nas áreas de especificação do modelo de negócio e validação do mesmo, o projeto não alcançaria os resultados obtidos.

As tecnologias escolhidas mostraram-se excelentes ferramentas para soluções técnicas requeridas no projeto. Todas as expectativas de performance foram atendidas, respeitando os recursos alocados ao projeto.

5.1 Sugestões para trabalhos futuros

Tomando como base o que foi desenvolvido e levando em consideração o amplo aceite que a ferramenta desenvolvida teve junto às mais de 150 lojas da rede de franquias, enxerga-se algumas oportunidades de melhoria no projeto desenvolvido.

Entre estas melhorias, destacam-se:

1. Implementação de um sistema de rastreamento dos relatórios distribuídos.
2. Disponibilização sob demanda dos relatórios atualizados.

Entre os pontos citados, o primeiro deles trará um ganho direto para a franqueadora, empresa a qual contratou este projeto. O ganho que será obtido através da implementação de um sistema de rastreamento dos relatórios distribuídos é dado pelo fato de que, atualmente, os

relatórios são gerados e distribuídos como uma via única de informação. Ou seja, a franqueadora distribui os relatórios prontos para serem utilizados pelas lojas via e-mail, mas não possui uma maneira eficiente de medir quais lojas de fato estão utilizando os relatórios. Bem como, não consegue mensurar como as lojas que usufruem da ferramenta disponibilizada estão consumindo a informação.

Fazendo uma analogia com a teoria de controle, é como se o projeto desenvolvido estivesse operando em malha aberta. Ou seja, existe uma entrada, um sistema que possui uma saída de acordo com a entrada, mas não existe uma realimentação da saída, melhorando a eficácia do controle.

Já o segundo dos pontos de melhoria citados, trará um ganho para os franqueados que estão recebendo os relatórios. Isto ocorrerá, pois, atualmente, para evitar que os lojistas recebam demasiados e-mails, são enviados relatórios atualizados somente uma vez por semana. Ao receber este relatório com as oportunidades atualizada da loja, os gestores da loja se reúnem com equipe de vendas para discutir plano de ação da semana. Porém, sabe-se que em um universo de 151 lojas, nem todos os gestores fazem a gestão da mesma forma e com a mesma frequência.

Sendo assim, para tornar o intervalo de disponibilização dos relatórios o mais assertivo possível, idealmente o próprio gestor da loja deveria conseguir obter relatórios atualizados. Desta forma não se envia relatórios demais para quem faz a gestão de oportunidades semanal, e não se envia menos relatórios do que o necessário para os gestores que fazem acompanhamento das oportunidades da loja mais do que uma vez por semana.

6 REFERÊNCIAS

- 1 FORBES. **Big Data: 20 Mind-Boggling Facts Everyone Must Read.** Forbes. Disponível em: <<https://www.forbes.com/sites/bernardmart/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/>>. Acesso em: Maio 2018.
- 2 HUXLEY, A. **Data Don't Cease To Exist.** Forbes, 2016. Disponível em: <<https://www.forbes.com/quotes/234/>>. Acesso em: 01 jun. 2018.
- 3 GASKELL, A. **Becoming a Data Driven Organization.** Forbes. Disponível em: <<https://www.forbes.com/sites/adigaskell/2016/10/28/becoming-a-data-driven-organization>>. Acesso em: 10 maio 2018.
- 4 JOHNSON, C. **A Dieta da Informação.** 1. ed. Brasil: Novatec, 2012.
- 5 EVELSON, B.; , N. N. **Business Intelligence – An Information Workplace Report.** [S.l.]: Forrester, 2008.
- 6 GARTNER. **Magic Quadrant for Analytics and Business Intelligence Platforms.** Lund. 2018.
- 7 SMART QLIK. **Como criar um dashboard eficiente?** Smart Qlik. Disponível em: <<http://www.smartqlik.com.br/como-criar-um-dashboard-eficiente/>>. Acesso em: 26 abr. 2018.
- 8 PHONOWAY. **Qual é o bem mais precioso da sua empresa?** Phonoway. Disponível em: <<http://www.phonoway.com.br/blog/item/219-qual-e-o-bem-mais-precioso-da-sua-empresa>>. Acesso em: 16 fev. 2018.
- 9 DUARTE, E. **SQL e Programação de Banco de Dados.**, out. 2006.
- 10 DATE, C. **SQL e Teoria Relacional.** 1. ed. São Paulo: Novatec, 2015.
- 11 TEOREY, T. **Projeto e Modelagem de Banco de Dados.** 2. ed. São Paulo: Elsevier, 2013.
- 12 OZIMAR, H. **Tabela Fato x Tabela Dimensão.** Microsoft, 2012. Disponível em: <<https://social.technet.microsoft.com/wiki/pt-br/contents/articles/12577.tabela-fato-x-tabela-dimensao.aspx>>. Acesso em: 13 abr. 2018.
- 13 COUGO, P. **Modelagem Conceitual e Projeto de Banco de Dados.** 1. ed. São Paulo: Elsevier, 1997.
- 14 PRATTE, D. **Use this architecture to structure your business intelligence solutions.**, Tech Republic, 2001.

15 SIRIUS. **Tela ERP**. Sirius. Disponível em: <<http://www.sirius.inf.br/wp-content/uploads/2011/10/Pedidos-visao-especifica.png>>. Acesso em: 21 maio 2018.

16 MUNZERT, S. **Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining**. 1. ed. London: Wiley, 2015.

17 BAZILIO, C. **Normalização de Esquemas de Banco de Dados**. , UFF (Universidade Federal Fluminense), 2009.

18 RAGOZIN, A. **Snowflake Schema**. Disponível em: <<http://blog.ragozin.info/2011/04/data-grid-pattern-snowflake-data-schema.html>>. Acesso em: 19 abr. 2018.

19 TONIAL, F. **Aposlita de QlikView Avançado**. 1. ed. [S.l.]: Toccato, 2014.

20 NEVES, M. A. O. **Indicadores de desempenho em logística**. In: _____ **Indicadores de desempenho em logística**. 1. ed. São Paulo: Mundo, 2009.

21 ALE. **Painel do Carro**. Ale, 2007. Disponível em: <<https://www.blogdaale.com.br/voce-entende-o-painel-de-controle-do-seu-carro/>>. Acesso em: 15 jun. 2018.